FUNCTIONAL ANALYSIS OF THE *MOSPD* GENE FAMILY

KATRIN BUERGER

THESIS PRESENTED FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
UNIVERSITY OF EDINBURGH
2010

To my family, future and present .

I declare that the work presented in this thesis is my own, unless otherwise stated.

Katrin Buerger

Acknowledgements

I would like to thank my supervisor Lesley Forrester, for all the support, guidance and most importantly encouragement throughout my PhD studies. Many thanks also go to the members of my PhD committee, Alistair Watt, Matt Sharp and Matt Bailey for their valuable feedback and advice. Furthermore, I thank the Wellcome Trust and Professor John Mullins for awarding me a 4-year PhD studentship. Special thanks go to all my colleagues, past and present, at the John Hughes Bennett Laboratory for their help and cheerful company. A special thanks also to Christine, for keeping me going through the writing of my thesis. Finally, I would like to thank Hamood for all his love, patience and support.

ABSTRACT

Mospd3, a gene located on mouse chromosome 5, was identified in a gene trap screen in ES cells. The gene trap vector integration in multiple copies into the putative promoter of the gene, resulted in a loss of expression of *Mospd3* at the trapped allele. In mice generated from ES cells carrying the vector integration it was found that the lack of *Mospd3* expression resulted in the death of a proportion of the homozygote mutants within the first day after birth. Homozygote neonates exhibited a thinning of the right ventricular free heart wall which resembles other mouse mutant phenotypes as well as human congenital heart defects caused by a loss of desmosome and adherens junction mediated cell adhesion between cardiomyocytes. The protein encoded by *Mospd3*, contains an N-terminal Major Sperm Protein (MSP) domain implicated as a mediator of protein- protein interactions, as well as a two C-terminal transmembrane domains.

Both, protein structure and phenotypic similarities with defects in desmosomal and adherens junction proteins suggests that Mospd proteins might play a role in cell adhesion and maintaining the structural integrity of the heart.

The phenotype of *Mospd3* mutants was highly dependent on genetic background, which led us to speculate that there might be genetic redundancy between *Mospd3* and its closest family member the X-linked *Mospd1*.

The aims of this thesis were to generate tools to better understand the function of the *Mospd* gene family in cardiac development as well as assessing genetic redundancy between *Mospd1* and *Mospd3*. A conditional gene targeting strategy was designed for both *Mospd* genes. Large genomic regions of the *Mospd1* and *Mospd3* loci were subcloned from bacterial artificial chromosomes (BACs) and using a recombineering approach, *loxP* sites and a drug selection cassette (neomycin) were placed in precise locations surrounding the MSP domain of both genes. The conditional targeting vectors were electroporated into both CGR8 and E14 ES cells and homologous recombinant clones were identified at a frequency of 2% and 1.3% for Mospd1 and Mospd3 respectively. Five euploid targeted clones for both *Mospd1* and *Mospd3* have been generated.

Transient expression of Cre recombinase in ES cells carrying the conditional *Mospd1* allele was used to delete the one copy of this X-linked gene. Phenotypic characterisation of this null ES cell line revealed that *Mospd1* is neither essential for ES cell viability and self-renewal, nor for the early differentiation of these cells towards a cardiac fate.

In order to investigate the mechanism of action of Mospd proteins, specific polyclonal antibodies were generated to detect either Mospd1 or Mospd3. These antibodies were purified and tested by western blotting using COS7 cells overexpressing either Mospd protein as well as mouse tissue lysates. Whilst the antibodies were found to detect the proteins and differentiate between Mospd1 and Mospd3, they showed insufficient purification to be used in co-localisation and co-immunoprecipitation experiments to identify interacting proteins and determine whether Mospd proteins are involved in cell adhesion complexes.

Monoclonal antibodies were subsequently generated and initial western blotting experiments showed promising results, indicating that these antibodies may be better suited for immunohistochemical analysis of Mospd proteins.

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

ABBREVIATIONS

α	anti- (in conjunction with antibodies i.e. αMospd1 and αMospd3)
A	atria
AJ	adherens junction
Alp	α-actinin associated LIM-domain protein
Ao	aorta
AP axis	anterior-posterior axis
ARVC	arrhythmogenic right ventricular myopathy
AS	aortic sac
AVV	atrioventricular valve
BAC	bacterial artificial chromosome
BACPAC	bacterial artificial chromosome/P1-derived artificial chromosome
β-gal	β-galactosidase
β-geo	β-galactosidase and neomycine-phosphotrasferase fusion construct
bp	base pair
BMP	bone morphogenic protein
BSA	bovine serum albumin
ºC	degrees Celsius
cDNA	complementary DNA
C.elegans	*Caenorhabditis elegans*
CGR8	a 129/Ola mouse strain derived ES cell line
CHORI	Children's Hospital Oakland Research Institute
Chr	chromosomes
cko	conditional knockout
CNC	cardiac neural crest
CNS	non-coding sequences

COS7	a monkey kidney cell line
CRAL-TRIO	cellular retinaldehyde-binding/triple function domain
Cre	Cre recombinase
CT	conotruncus
DA	ductus arteriosus
DCM	dilated cardiomyopathy
DH5α	*E.coli* strain
DH10B	*E.coli* strain
DMSO	dimethyl-sulfoxide
DNA	deoxyribonucleic acid
DNA	deoxyribonucleic acid-transposon (only used in Figure 3.1 and 3.6)
dNTP	deoxyribonucleotide triphosphate
ds	double-stranded
E	embryonic day
E	enhancer (only used in Figure 1.8)
EB	embryoid body
ECM	extracellular matrix
E. coli	*Escherischia coli*
eGFP	enhanced green fluorescent protein
E14	E14tg2a- a 129/Ola mouse strain derived ES cell line
EB	embryoid body
ELISA	enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
ER^{T2}	tamoxifen inducible estrogen receptor
ES	embryonic stem cell
EUCOMM	European Conditional Mouse Mutagenesis project

FGF	fibroblast growth factor
FHF	first heart field
FIAU	Fialuridine
FISH	fluorescent *in situ* hybridisation
FlEX	flip-excision
f*lox*ed	flanked by *loxP* sites
Flpe	enhanced flipase
for	forward
foxh1	forkhead box H1
FRT	flipase recognition site
G418	geneticin antibiotic (resistance conferred by *neo* gene)
Gata	GATA binding protein
GEO	Gene Expression Omnibus
GLVP	Gal4/ herpes simplex virus VP16 transcriptional transactivation domain and human truncated progesterone receptor fusion
h	hour
Hand	heart and neural crest derivatives expressed transcript
HAT	hypoxanthine aminopterin thymidine medium
HPRT	hypoxanthine-guanine phosphoribosyltransferase
HSV-tk	herpes simplex virus thymidine kinase
hygro	hygromycin B - phosphotransferase
ICD	intercalated disc
ICM	inner cell mass
Ig	immunoglobulin
IGTC	International Gene Trap Consortium
it	incomplete targeted allele

kb	kilo base
kDa	kilo Dalton
KLH	keyhole limpet heamocyanin carrier protein
KO	knockout
KOMP	Knock-Out Mouse Project
KSCN	potassium thiocyanate
Lamp2	lysosomal-associated membrane protein 2
LB	Luria-Bertani bacterial culture broth
LBD	ligand binding domain
LIF	leukaemia inhibitory factor
LIM	Lin11, Isl-1 and Mec-3 domain
LIN	long interspersed nuclear elements
loxP	locus of chromosomal crossover P1
LSA	left subclavian artery
LTR	long terminal repeats
LV	left ventricle
M	mol/molar
mCi	millicurie
Mef2c	myocyte enhance factor 2c
MerCreMer	*Cre* recombinase gene flanked by two murine estrogen receptors
µF	micro Faraday
MF1	mouse strain with mixed genetic background
MHC	myosin heavy chain
min	minute
Mlc2υ	myosin light chain 2υ
Mlp	muscle specific LIM-domain protein

Mospd	motile sperm domain-protein
Mospd3^{gt1lmf}	*Mospd3* gene trap allele 1, generated in Lesley Forresters' laboratory
MSP	major sperm protein/domain
N	mouse breeding generation number
N-cadherin	neuronal cadherin
NCBI	National Center for Biotechnology Information
neo	neomycin phosphotransferase gene
neo^{R}	resistance to neomycin
Nkx2.5	Nk2 transcription factor related 5
NLS	nuclear localisation sequence
NorCOMM	North American Conditional Mouse Mutagenesis project
Nppa	natriuretic peptide precursor type A
Ω	Ohm
O.D.	optical density
OFT	outflow tract
P	promoter
PA	pulmonary artery
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDZ	post synaptic density protein, Drosophila disc large tumor suppressor and zonula occludens-1protein
Pfam	protein family database
Pgk	phosphoglycerate kinase 1
Pitx2	paired-like homeodomain transcription factor 2
polyA	polyadenylation
puro	puromycin

qPCR	quantitative polymerase chain reaction
RA	retinoic acid
RA	right artery (only used in Figure 1.4)
RCA	right carotid artery
rev	reverse
RNA	ribonucleic acid
RNA	ribonucleic acid retrotransposon (only used in Figures 3.1 and 3.6)
RNAi	ribonucleic acid interference
rpm	revolutions per minute
RSA	right subclavian artery
RT	room temperature
RTs	recombinase target sites (only used in Figure 1.10)
rtTA	reverse tet-on tetracycline controlled transactivator
RV	right ventricle
RVW	right ventricular wall
SA	splice acceptor
SD	splice donor
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
sec	second
SHF	second heart field
SIN	short interspersed nuclear elements
SOC	superoptimal culture broth
ss	single-stranded
SSC	saline-sodium citrate buffer
t	targeted allele, mutation or ES cell line

TBS	tris-buffered saline buffer
TE	Tris-EDTA buffer
TGFβ	transforming growth factor β
TM	transmembrane domain
TSP	tissue-specific promoter
tTA	tetracycline controlled transactivator
UTR	untranslated region
V	ventricle (only used in Figure 1.4)
V	volts
VAP	vesicle associated membrane protein (VAMP)-associated protein
wt	wild type
wt/vol	weight per volume

CHAPTER 1: INTRODUCTION

1.1 WHY STUDY MOSPD GENES?

Motile sperm domain-protein (*Mospd*) *3* a member of the *Mospd* gene family was identified in a gene trap screen designed to identify genes which are involved in embryonic development (Forrester et al., 1996; McClive et al., 1998). Transgenic mice carrying the gene trap integration showed neonatal death in a proportion of homozygote animals as well as cardiac dilation in surviving litter mates of that genotype. The phenotype suggested that Mospd3 might be involved in the development or function of the heart (Pall et al., 2004).

In studying the role of *Mospd3* during development we hope to gain not only information regarding the function of this largely uncharacterized family of genes but also new insights into cardiac development and function.

1.1.1 The Motile Sperm Domain-protein Family

Mospd3 encodes a 235 amino acid protein, which belongs to a family of 4 proteins in the mouse. Mospd1 to 4 are all defined by the presence of a major sperm protein (MSP) domain. In addition, all four proteins also contain one or two C-terminal transmembrane domains (Figure 1.1 and 1.2).

Protein sequence alignments (using ClustalW[1]) as well as analysis of the structural domains between Mospd proteins in the mouse revealed that Mospd1 shared the greatest amount of protein sequence and structural identity with Mospd3, whilst both Mospd2 and 4 are evolutionary more diverged (Figure 1.1). Even though the protein conservation between Mospd1 and Mospd3 may be as low as 38%, most of the sequence identity is found in the MSP and transmembrane domains (Figure 1.2).

Mospd3 appears to be mammalian specific and is 88% similar to its human orthologue encoded on Chromosome 7q22, a region syntenic to chromosome 5 in mouse (Pall et al., 2004). In contrast to *Mospd3, Mospd1* is conserved in a wide range of vertebrate organisms: from zebrafish, *Fugu* and *Xenopus*; to chicken, rat and human. *Mospd1* is encoded on the X-chromosome in mouse and human.

[1] http://www.ebi.ac.uk/Tools/clustalw2/index.html

Interestingly, rats possess two *Mospd1* genes: one located on the X-chromosome, whilst its paralogue maps to rat chromosome 14 (Pall et al., 2004).

Expression analysis and data obtained from the Gene Expression Omnibus (GEO) database[2] show that both *Mospd1* and *Mospd3* are expressed in a wide range of embryonic and adult tissues; indicating that their expression may be ubiquitous. The widespread expression of both genes was confirmed in RT-PCR and *in situ* studies. *In situ* analysis of *Mospd3* detected expression of the gene throughout the late gestation embryo (Pall et al., 2004). Similarly, RT-PCR analysis on a range of available test cDNA samples including those of hematopoietic tissues in the embryo (fetal liver, yolk sac and aorta-gonad-mesonephros tissue) showed expression in all these tissues as well as in *in vitro* generated embryoid bodies at day 3 and day 5 (explanation of *in vitro* differentiation of ES cell in section 1.4.2). *Mospd1* also showed expression in the hematopoietic tissues and embryoid bodies (Forrester, unpublished). This overlapping expression pattern of both *Mospd1* and *Mospd3* could indicate a functional redundancy between these two genes. In addition to these embryonic tissues *Mospd3* was found to be expressed in neonatal brain, kidney, liver, spleen and heart (Pall et al., 2004). Lastly, both *Mospd3* and *Mospd1* were also shown to be expressed in adult heart (Forrester, unpublished). In order to gain insight into the function of these *Mospd* genes more detailed expression analysis throughout mouse development may be required, and perhaps more importantly information about the expression and localisation of the protein products encoded by both genes is needed.

1.1.2 Function of the Major Sperm Protein domain

A study looking at MSP-domain proteins identified 199 members in the Pfam database (Tarr and Scott, 2005). The majority (about 60%) of this group of proteins was made up of nematode sperm-specific major sperm protein. The first of these to be identified was the major sperm protein in *C. elegans* (Burke and Ward, 1983; Scott et al., 1989).

[2] http://*www*.ncbi.nlm.nih.gov/geo/

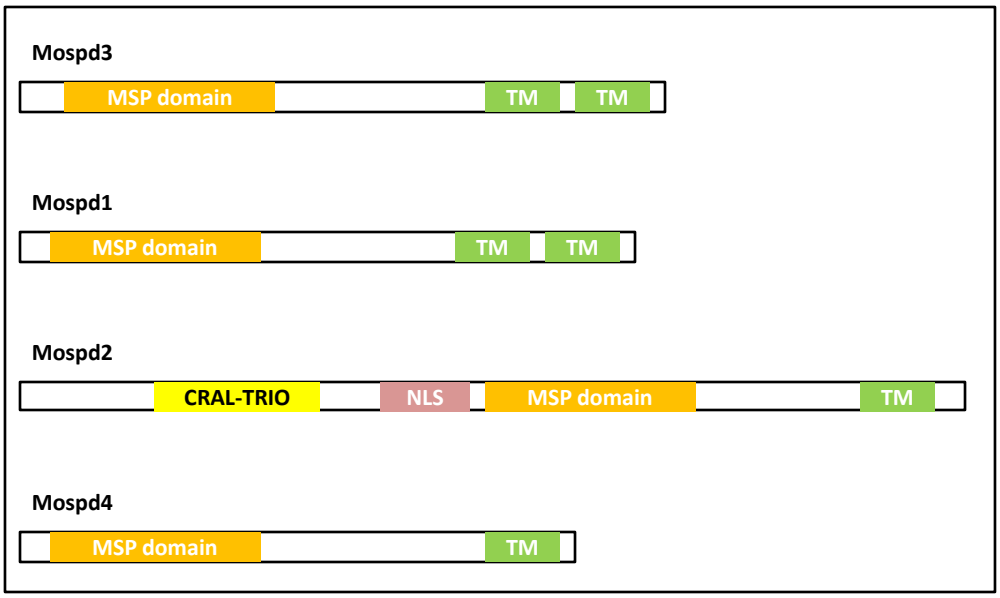


Figure 1.1: Mospd protein family in *Mus musculus*. In mouse this family contains four members, which all contain a Major Sperm Protein domain (MSP). Mospd1 and Mospd3 proteins show the highest degree of protein structure conservation with both containing two C-terminal transmembrane (TM) domains, whilst Mospd2 and Mospd4 only contain only one transmembrane domain. Mospd2 also includes a nuclear localisation (NLS) domain and a Cellular retinaldehyde-binding/triple function (CRAL-TRIO) domain.

Figure 1.2: Protein sequence alignment of Mospd3 and Mospd1 using (ClustalW). Although the global alignment score between the to proteins is only 38% most of the sequence similarity is contained within the major sperm domain marked with an orange box and the two C-terminal transmembrane domains marked with green boxes. The colouring of the amino acid residues identifies their physiochemical properties. Red indicates small hydrophobic amino acids which include aromatic residues; blue marks acidic residues; magenta marks basic residues; green indicates residues containing hydroxy-,sulfhydryl or amine groups.
Asterisk (*), colon (:) and single dot (.) refer to the level of conservation between the amino acid residues in Mospd1 and Mospd3. The asterisk (*) indicates identical amino acid residues between the two protein sequences. The colon (:) marks amino acid substitution with the same physiochemical properties, whilst the single dot (.) indicates semi-conserved amino acid substitutions.

One of the functions of MSP (major sperm protein) in nematodes is as a signalling molecule to promote oocyte maturation and sheat cell contraction in female animals (Miller et al., 2001). More importantly, MSP acts as a nematode-specific cytoskeletal protein; required to allow crawling motility in nematode sperm (Italiano et al., 2001; King et al., 1994b). Blocking of major sperm protein function in nematodes leads to male infertility.

Analysis of the crystal structure of nematode major sperm protein revealed that the protein assumes an immunoglobulin (Ig)-like fold structure (Bullock et al., 1996). This fold is also found in a variety of other proteins such as fibronectin, laminin and human growth factor. The Ig-like domain interacts with other Ig-like domains via the β-sheet region of its structure. This interaction forms the basis of the assembly of MSP monomers into dimers; and further into helical filaments and higher order structures called macrofibres (King et al., 1994b). MSP filaments assemble at the leading edge of a pseudopod formed by nematode sperm. This assembly mechanism drives motility in these cells (King et al., 1994b; Roberts and King, 1991).

It has been noted that the motility function, which major sperm protein provides in nematode sperm, is almost indistinguishable from the actin-based motility observed in other cells (Roberts and Stewart, 2000).

1.1.2.1 Comparison of MSP and actin based motility

Actin facilitates motility in a variety of cells such as amoeba and white blood cells. Movement of these cells is achieved through the formation of an actin-rich pseudopod structure at the front of these cells. The pseudopod attaches to the substrate and pulls the cells along the surface. Sperm movement in nematodes utilises a similar pseudopod structure. However, as nematode sperm do not contain significant amounts of actin (Nelson et al., 1982); the locomotion of these cells is based on the MSP protein in the pseudopod.

Akin to actin-based motility; sperm locomotion functions by MSP filament assembly at the leading edge of the pseudopod and disassembly of the filaments near the cell body at the base of the pseudopod structure (King et al., 1994b; Roberts and King, 1991). As a result of this treadmilling process the cell is pulled along.

The assembly and disassembly of the actin cytoskeleton in the pseudopod requires ATP hydrolysis as well as a whole host of accessory factors; to orchestrate this process (Borisy and Svitkina, 2000; Pollard and Borisy, 2003). In contrast, MSP filaments appear to be able to polymerise spontaneously without the need for ATP binding and hydrolysis (Italiano et al., 1996). Instead, MSP filament assembly and disassembly in the pseudopod is partly due to a pH gradient of 0.2 units higher at the leading edge compared to the cell body end (King et al., 1994a).

More recently, it was found that similar to the actin cytoskeleton; the MSP cytoskeleton also employs a number of accessory factors to modulate MSP filament assembly. These factors include MFP1, MFP2 (MSP fibre protein) and MPOP (MSP polymerization organizing protein). Whilst MPOP is instrumental in directing the assembly of MSP filaments at the membrane of the pseudopod's leading edge (LeClaire et al., 2003); MFP1 and MFP2 have been found to modulate the growth of MSP filaments (Buttery et al., 2003; Grant et al., 2005).

MSP and actin-based motility appear to be morphologically and dynamically analogous despite being based on entirely different sets of proteins (Roberts and Stewart, 2000). Although the proteins involved in either of these two cytoskeletal systems seem to be mutually exclusive; a study looking at nematode sperm-specific clusters of gene transcripts has identified increased levels of expression of 3 clusters of ESTs encoding proteins, containing either a LIM or a PDZ domain (Tarr and Scott, 2004). Both of these protein-protein interaction domains have been associated with the actin cytoskeleton and its organization (Khurana et al., 2002; Pomies et al., 1999; Xia et al., 1997). LIM domains are found in a variety of different proteins. Whilst LIM-domain proteins found in the nucleus function in tissue-specific gene regulation and cell fate determination; cytoplasmic LIM-domain proteins have been shown to facilitate actin filament organization (Zheng and Zhao, 2007). A number of these cytoplasmic LIM-domain proteins also contain a PDZ domain. This group of proteins includes Alp and Mlp; two proteins which are further discussed in section 1.3.2.

The PDZ protein-protein interaction domain functions by facilitating the formation of branched networks of target proteins via specific C-terminal binding sequences (Harris and Lim, 2001). PDZ proteins, like Alp and Mlp, mediate actin

bundling by binding through the PDZ domain to the spectrin-like motifs of α-actinin; thereby enhancing α-actinin-associated actin cytoskeleton organization (Pomies et al., 1999; Xia et al., 1997). As many of the MSP domain encoding genes also code for potential PDZ domain binding sites, it has been speculated that PDZ-domain proteins may also modulate MSP cytoskeleton assembly (Tarr and Scott, 2004). The possibility of an interaction of PDZ-domain proteins with both actin and MSP filaments could indicate a link between these two different and distinct cytoskeletal systems.

1.1.3 Major Sperm Protein domain proteins in other eukaryotic organisms

Apart from the major sperm protein in nematode, very little is known about the function of MSP-domain proteins. The only functional information for this type of protein in other model systems comes from the study of vesicle associated membrane protein (VAMP)-associated proteins (VAP) proteins.

Similar to Mospd proteins, VAP proteins encode an N-terminal MSP domain as well as a C-terminal transmembrane domain. Both these domains are separated by a coiled-coil domain not found in Mospd. VAP proteins have been described in vertebrates (Nishimura et al., 2004; Skehel et al., 2000); and lower organisms such as *Drosophila* (Pennetta et al., 2002), *Aplysia* (Skehel et al., 1995), yeast (Kagiwada et al., 1998) and plants (Laurent et al., 2000).

The first of these proteins to be identified was the VAP33 protein of *Aplysia*, which was found to be involved in synaptic transmission in the central nervous system (Skehel et al., 1995). The next VAP to be described was in *Drosophila.* Interestingly, two *VAP* genes are expressed in this system. Whilst one of the protein products is required for male fertility; the protein of the other is found to localise at the neuromuscular junction; where it appears to be involved in the transport of vesicles towards the synaptic membrane (Pennetta et al., 2002). The mouse orthologue, *VAP33,* was shown to serve a similar vesicle trafficking function; a finding supported by the localisation of mouse VAP33 protein to: not only the endoplasmic reticulum but also the microtubules and intracellular vesicles (Skehel et al., 2000). A mutation in the *VAP-B* gene in humans was recently described

(Nishimura et al., 2004). Affected individuals suffered from amyotrophic lateral sclerosis - a motor neuron disease. Disease in these individuals is the result of a point mutation in the sequence encoding the MSP domain of VAP-B, which causes a conformational change in the Ig-like domain. The mutant VAP-B protein fails to localise to endoplasmic reticulum (ER), vesicles or microtubules (Mitne-Neto et al., 2007). Evidence gained from these VAP protein studies indicates that MSP-domain proteins may be functioning as mediators of protein-protein interaction within or between cells.

1.1.4 Phenotype of *Mospd3*$^{Gt1lmf/Gt1lmf}$ mice

Besides the comparative analysis with other MSP-domain proteins, the only other clue to the function of Mospd proteins has come from a mouse mutant generated by gene trapping (called R124 gene trap integration) that appeared to result in a null allele of *Mospd3* (McClive et al., 1998; Pall et al., 2004).

When mice carrying the *Mospd3*^{Gt1lmf} allele were generated on a C57BL/6 inbred genetic background, a reduction in the number of expected homozygous mutant mice was observed (Pall et al., 2004). It was subsequently discovered, by close monitoring of the homozygotes obtained by caesarean section at embryonic day 19.5 (E19.5) compared to those taken after natural birth, that about half of the homozygous pups died within the first 24 hours of their neonatal life. This was not the case when the gene trap integration was crossed onto an outbred MF1 background. On this mixed genetic background, even after three weeks, the normal Mendelian ratio of homozygote pups was observed; and homozygous mice did not display any obvious phenotypic defect. As the phenotype appears dependent on genetic background, and even in the inbred strain only half the homozygote offspring die, this would indicate the presence of a genetic modifier.

On the C57BL/6 inbred background, the gene trap integration leads to a defect in the morphology of the cardiac right ventricle in homozygous pups delivered just before birth by caesarean section. Heterozygote animals carrying only one trapped allele did not display this defect. When the hearts of homozygous pups were examined, they were found to display dilation of the right ventricle with a pronounced thinning of the right ventricular parietal (free) wall as well as an

attenuation of the growth of the papillary muscle and supraventricular crest (Figure 1.3). The ventricular region near the apex of the heart was most affected by this thinning of the heart wall. Measurements of this region in homozygotes, compared to unaffected heterozygous and wild type littermates, showed a reduction of 37% in the thickness. In addition, the volume of the right ventricular chamber was found to be increased by 56%. The observed thinning of the right ventricular wall appears to be the result of a narrowing of the compact myocardium and a reduction in the number and size of the trabeculae carnae.

The hearts of *Mospd3*$^{Gt1lmf/Gt1lmf}$ embryos were also analysed at E16.5 (Forrester, unpublished data). The heart, at this stage, does not differ from that of heterozygote and wild type littermates, indicating that the right ventricular defect is associated with the growth, maturation or structural integrity of the ventricle between E16.5 and birth.

The phenotype observed in *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals shows that Mospd3 may be required for the proper development of the heart, in particular the right ventricle. The observation that right ventricular thinning does not arise until very late in gestation suggests that the defect may be caused by problems in the maturation of the right ventricular compartment. An alternative explanation may be that Mospd3 is involved in providing structural integrity to the ventricle; and in its absence the developing heart fails to cope with the ever-increasing hemodynamic forces, it is subjected to, during later embryonic development; and after birth when the hemodynamic load increases exponentially.

In addition to the neonatal death phenotype observed in *Mospd3*$^{Gt1lfm/Gt1lfm}$ mice, an adult phenotype of hypertrophy of the heart, kidney and testis has been described in these homozygote mutant animals (Pall, 1998). However, this phenotype was only observed during early backcrosses (N2-4) of the gene trap integration into a pure C57BL/6 genetic background, and had been lost by generation N10. This adult phenotype may have been associated with the particular mixed genetic mouse background of 129/Ola and C57BL/6 strains, or with an unrelated genetic mutation which had occurred during the gene trap integration. In both cases, further crossing onto a C57BL/6 background would have led to the loss of this phenotype.

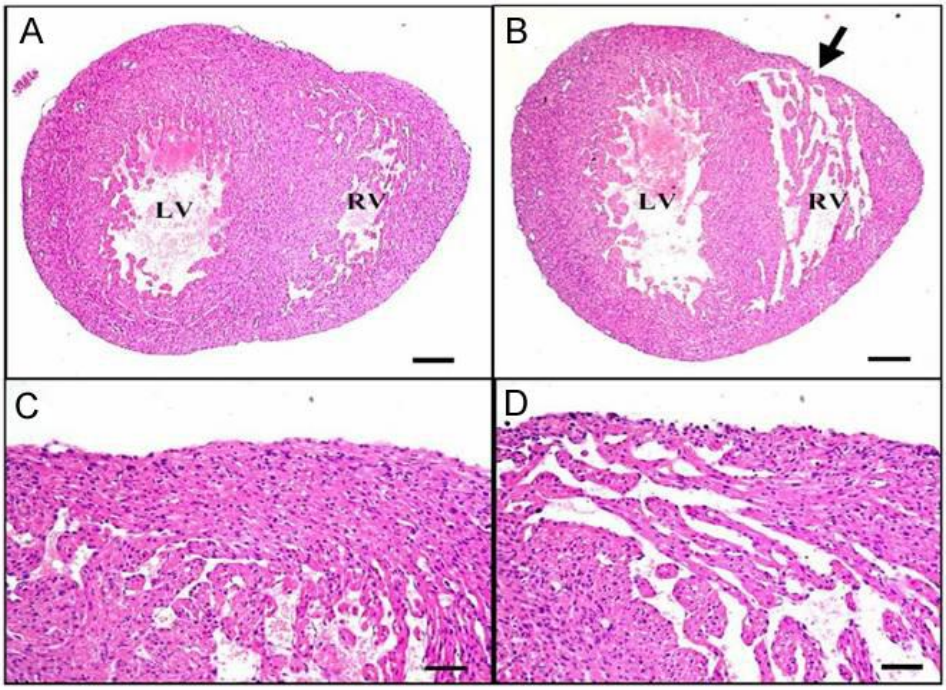


Figure 1.3: Transverse heart sections from heterozygous *Mospd3*$^{Gt1lmf/+}$ (A and C) and homozygous *Mospd3*$^{Gt1lfm/Gt1lfm}$ (B and D) embryos at E19.5. Haematoxylin and eosin stained cardiac tissue sections revealed that compared to the heterozygous heart (A) the compact myocardium of the right ventricular free heart wall in the homozygote heart (B) shows pronounced thinning. The layer of compact myocardium is thinnest in the anterior region of the right ventricle of the homozygous heart marked by the arrow in (B) and shown at greater magnification in (D). (C) Magnified view of the anterior right ventricle of the section shown in (A) through the heart of a heterozygote embryo. The homozygote heart (A and C) also shows a reduction in the number of trabeculae carnae compared to the heterozygote heart (B and D). Abbreviations: LV, left ventricle: RV, right ventricle. Bar (A and B) = 100µm and bar (C and D) = 25µm. Figure adapted from Pall *et al*. (2004)

Interestingly, hypertrophy of the heart - specifically the right ventricular wall- was also observed in later generations (beyond N10), when a litter of *Mospd3*$^{Gt1lfm/Gt1lfm}$ neonates along with their heterozygous littermates was retrieved within the first 24 hours after natural birth, and analysed for heart morphology (Forrester, unpublished). One of the homozygote neonates had died before being taken for analysis. Morphological analysis of the hearts of these animals performed by Dave Brownstein[3] showed that all the homozygote animals, including the one which had died, displayed a hypertrophy of the septum transversum and the right ventricular free heart wall. The tissue thickening was primarily observed in the compact cell layer of the affected heart wall. It is possible that lack of Mospd3 in the heart could result in a loss of structural integrity of the heart muscle or reduced contractile force transmission in the ventricle (described in more detail in section 1.3). The observed postnatal right ventricular wall hypertrophy may be an adaptation mechanism of the cardiac muscle, in order to cope with the high hemodynamic pressures and to compensate for decreased structural integrity. This observation was not analysed in further litters, as at this time the neonatal death phenotype in the *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals began to be lost. Nearly all the homozygote neonates carrying the gene trap integration, at this stage, survived into adulthood. This fact along with the complexities of the gene trap integration (described in detail in section 1.4.5.4) have left us to consider the generation of an alternative *in vivo* mouse model for the study of *Mospd3* (described in detail in Chapter 3).

[3] Mouse Pathologist at the Queen's Medical Research Institute, Edinburgh

1.2 HEART DEVELOPMENT

The contractile heart is the first fully functioning organ present in the embryo. Its morphogenesis, including alignment of the cardiac chambers and formation of vascular connections, is complete by late mid-gestation. The subsequent period of cardiac development involves the maturation and further growth of the structure.

Development of the heart starts during gastrulation, when cells from the anterior region of lateral mesoderm become committed to the cardiogenic fate. Although little understood, this commitment is thought to be achieved in part by signals from the adjacent endoderm along with morphogens such as bone morphogenic proteins (BMP) (Schultheiss et al., 1997), fibroblast growth factors (FGF) and Wnts (Marvin et al., 2001; Schneider and Mercola, 2001). Cardiac progenitor cells subsequently form a bilateral symmetric heart field (Figure 1.4 A), also called the cardiac crescent, which fuses at the ventral midline to form a beating linear heart tube (Figure 1.4 B). Factors such as Nkx2.5 (Nk2 transcription factor related 5), Gata (GATA binding protein) and Mef2c (myocyte enhancer factor 2C) have been implicated in this process (Srivastava and Olson, 2000). The linear heart tube, formed by E8, comprises an outer layer of myocardium and an inner layer of endocardium separated by extracellular matrix (ECM). Even as this tubular structure develops, it becomes patterned along its anterior-posterior (AP) axis, specifying the future structures of the four chambered heart – namely atria, left and right ventricle, conotruncus and the aortic sac (Redkar et al., 2001).

Initiated by the asymmetric axial expression of factors such as *TGFβ* (*transforming growth factor β*), *Nodal* and *Pitx2* (*paired-like homeodomain transcription factor 2*) in the embryo, the vertebrate heart undergoes rightward looping (Capdevila et al., 2000). This process is essential to establish the correct orientation of the right (pulmonary) and left (systemic) ventricle as well as aligning the heart chambers with their respective vascular connections. The morphogenesis of the heart culminates in the septation of the four cardiac chambers by dividing both atria, and the right and left ventricles, through septa and the formation of the two atrioventricular valves, which arise from the cardiac cushion. As a consequence of septation, the flow of blood in the embryo becomes partitioned into pulmonary and systemic circulation.

Finally, the connections from the heart to the vascular system are fashioned by neural crest cells from the pharyngeal arches invading the developing heart to help pattern the outflow tract and the great arteries (Farrell et al., 1999).

1.2.1 Specification of the heart – Two heart fields

As early as the cardiac crescent, cells are fated to contribute to specific cardiac compartments (Figure 1.4). Upon looping of the heart, the developing chambers become morphologically distinct and establish specific chamber identities, identifiable by the expression of different combinations of transcription factors as well as structural proteins such as myosin light chains (Small and Krieg, 2004). These physiological and gene expression pattern differences distinguish not only atrium from ventricle, but also between right and left ventricle.

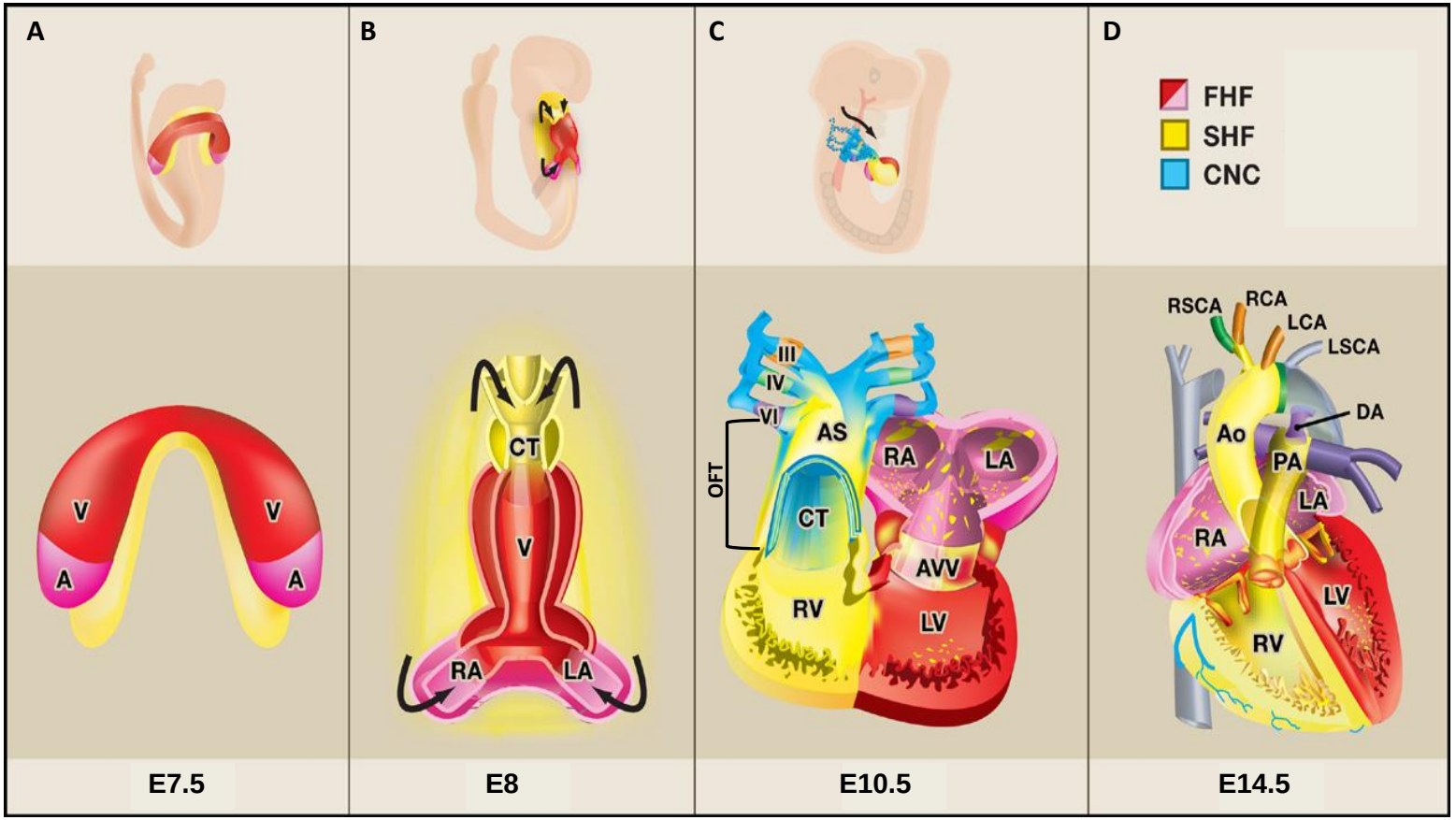
Recent lineage tracing studies in both the chick (Mjaatvedt et al., 2001; Waldo et al., 2001) and the mouse (Cai et al., 2003; Kelly et al., 2001) have revealed that these differences, particularly between the two ventricles, are due to their origin from two distinct populations of cardiac precursor cells (Figure 1.4). These studies have shown that a second population of progenitor cells, named the second or anterior heart field, contributes to the developing heart.

In the mouse, this second heart field arises from cells located anterior and dorsal to the linear heart tube, which itself is derived from the first heart field (Figure 1.4) (Cai et al., 2003). As the heart tube is formed, cells from the second heart field migrate into the anterior and posterior ends of this structure (Figure 1.4 B), where upon looping they contribute to the outflow tract and the right ventricle (Figure 1.4 C). The left ventricle and both atria are formed by cells of the first heart field; and the left ventricle, in particular, is only sparsely populated by progenitors from the second heart field (Figure 1.4 D) (Cai et al., 2003).

1.2.2 Morphogenesis of the right ventricle

In correspondence with their origin from two distinct progenitor populations, the right and left ventricles have also been shown, through mouse mutagenesis studies, to require different transcription factors for their proper morphogenesis.

Figure 1.4. Contribution of the first and second heart fields to the developing mouse heart. Oblique views of whole embryos (in upper panels) and frontal views of the cardiac structure and precursors (in lower panel) are shown. (A) First heart field (FHF) cells form a crescent shape in the anterior embryo at E7.5 with second heart field (SHF) cells medial and anterior to the first heart field. (B) Once the primitive heart tube has formed at the ventral midline of the embryo (B, upper panel) SHF cells positioned dorsal to the heart tube start to migrate (arrows) into the anterior and posterior ends of the tube (B, lower panel) to form the right ventricle (RV), the conotruncus (CT) and part of the atria (A). (C) Following the rightward looping of the heart tube, the right and left ventricular (RV and LV) chambers of the heart are formed and start to mature (C, lower panel). Also at this stage cardiac neural crest (CNC) cells migrate (arrow, C, upper panel) from the dorsal neural tube into the outflow tract (indicated by bracket; OFT) to form the vascular connections of the heart (marked in blue, the aortic arch arteries III, IV and VI). (D) The septation of the atria and ventricles results in the formation of the four chambered heart. Both atria and left ventricle of the heart are mainly derived from precursors of the first heart field (pink and red), whilst the right ventricle and outflow tract are almost exclusively formed from precursors of the second heart field (yellow). Abbreviations: A, atria; Ao, aorta AS, aortic sac; AVV, atrioventricular valve; CNC, cardiac neural crest; CT, conotruncus; DA, ductus arteriosus; FHF, first heart field; LA, left atrium; LCA, left carotid artery; LSA, left subclavian artery; LV, left ventricle; OFT, outflow tract; PA, pulmonary artery; RA, right atrium; ; RCA, right carotid artery; RSA, right subclavian artery; RV, right ventricle; SHF, second heart field; V, ventricle; aortic arch arteries numbered III, IV and VI. Figure adapted from Srivastava *et al*. (2006).


A
B
C
D
FHF
SHF
CNC
V
V
A
A
CT
V
RA
LA
III
IV
VI
AS
OFT
CT
RA
LA
AVV
RV
LV
RSCA
RCA
LCA
LSCA
DA
Ao
PA
LA
RA
LV
RV
E7.5
E8
E10.5
E14.5

One of the first markers to be expressed in the second heart field is *Islet1* (*Isl1*) (Cai et al., 2003). Its absence in mutant embryos appears to cause a defect in the expansion of the second heart field along with an inability of this cell population to migrate into the linear heart tube, resulting in the specification of only two cardiac chambers in these mice. Genetic marker analysis of the two cardiac chambers revealed that they correspond to the left ventricle and the atria, whilst the right ventricle failed to form. In addition, the absence of the outflow tract confirms the lack of contribution of the second heart field derivatives to the cardiac structure of *Isl1*$^{-/-}$ embryos (Cai et al., 2003).

Another transcription factor found to be required for the development of the right ventricle is Hand2 (heart and neural crest derivatives expressed transcript 2, formerly dHand) (Srivastava et al., 1995; Srivastava et al., 1997). Mice lacking Hand2 show right ventricular hypoplasia, indicating that cells from the second heart field fail to expand to form the right ventricle (Srivastava et al., 1997; Yamagishi et al., 2001). In contrast, Hand1 (heart and neural crest derivatives expressed transcript 1; formerly termed eHand) appears to be important for the development of the left ventricle. While *Hand2* is predominantly expressed in the right ventricle, *Hand1* expression is enriched in the outer curvature myocardium of the left ventricle (Yamagishi et al., 2001). *Hand1* is down-regulated in *Nkx2.5* deficient embryos that fail to properly specify the ventricular compartments of the linear heart tube and die around the time of cardiac looping (Biben and Harvey, 1997). Conditional deletion of *Hand1,* specifically in the heart, results in left ventricular hypoplasia (McFadden et al., 2005), indicating that Hand1 may be required for the proliferation of progenitors of the first (primary) heart field. Loss of both *Nkx2.5* and *Hand2* in a double knockout (Yamagishi et al., 2001), or compound loss of both *Hand* genes (McFadden et al., 2005), results in the complete absence or expansion failure of both ventricles, respectively.

Consistent with these findings in the mouse mutant model; studies in fish, which only develop one ventricle, show that lack of the single fish *Hand* orthologue, *hand2,* results in a failure to develop a ventricular chamber (Yelon et al., 2000).

Additional factors shown to be required for the development of second heart field derivatives - such as the right ventricle and outflow tract - are the transcription

factors Mef2c (myocyte enhancer factor 2C) (Lin et al., 1997) and Foxh1 (forkhead box H1) (von Both et al., 2004) as well as the chromatin remodelling protein Smyd (or Bop) (Gottlieb et al., 2002). All three of these genes have been implicated as upstream regulators of *Hand2* expression (Phan et al., 2005; Srivastava, 2006), and deletion of any of these genes was found to result in a phenotype resembling that of *Hand2* mutation in mice, leading to hypoplasia of the right ventricle and incomplete development of the outflow tract.

1.2.3 Heart Chamber Maturation

Initial specification and morphogenesis of the cardiac chambers of the looped heart tube is followed by the maturation of the respective heart chambers. This process is particularly important in the ventricles, as ventricles need to adapt to the increasing hemodynamic load of the developing embryo.

In ventricles this maturation process involves the expansion (or ballooning) of the chambers along with cardiomyocyte expansion and the formation of trabeculae. Only some of the factors involved in this process are known. Amongst these are the endocardial growth factor neuregulin and its receptors Erbb2 and 4 (v-erb-b2 erythroblastic leukaemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog). Mutant mice deficient in any of these three genes displayed absence of trabeculae, a slight reduction in the compact myocardium of the ventricles and endocardial cushion defects, leading to the death of embryos in mid-gestation (Garratt et al., 2003; Gassmann et al., 1995; Meyer and Birchmeier, 1995).

In addition to these endocardial factors, BMP10 and the polycomb protein gene BAF180 have recently been implicated in the proliferation of the ventricular myocardium (Chen et al., 2004; Wang et al., 2004). Mice lacking either gene display hypoplastic ventricular development. Ablation of BMP10, which is required for the maintenance of Nkx2.5 and Mef2C expression levels, also leads to absence of trabeculae and embryonic death at E10.5 (Chen et al., 2004).

1.3 STRUCTURAL INTEGRITY OF THE HEART AND CARDIOMYOPATHY

In addition to the expansion of the compact myocardium during heart chamber maturation, the establishment of structural integrity in the heart tissue, especially in the ventricular myocardium, represents another mechanism to cope with the increased hemodynamic load in the developing embryo and after birth. Structural integrity in cardiac muscle involves stable connections of neighbouring cardiomyocytes and the integration of the contractile muscle cytoskeleton at the centre of the cardiomyocytes with the cell membrane. As the hearts primary function consists of continuously pumping blood, in order to supply the whole body with oxygen and nutrients, an efficient system of force transmission from the contractile structure of the cardiomyocyte to the rest of the cell as well as amongst all the cardiac muscle cells in the heart is essential.

The contractile force in cardiac muscle is generated by myofibril structures, which fill nearly the entire mature rod-shaped cardiomyocyte cell. These myofibrils span the whole length of the cardiomyocyte and are attached at both ends to the intercalated disc, a structure at the polar membranes of cardiomyocytes (Figure 1.5). Each myofibril is made up of sarcomere subfilaments connecting to each other *via* their Z-disc structures, at which the actin filaments of each sarcomere are anchored. The muscle contraction in cardiomyocytes is a result of the interaction of actin (also thin filaments) with myosin filaments (thick filaments) in the sarcomere.

The contractile force generated by the sarcomeres in cardiac, and all other kinds of striated muscle, is transmitted to other organelles of the muscle cell and to the cell membrane though non-sarcomeric cytoskeletal components, which tether and maintain the contractile apparatus. Long filamentous actin fibres at the cell periphery tether the myofibrils *via* their Z-discs to adherens junctions of the cardiomyocyte intercalated disc. Another cytoskeletal component helping to integrate the contractile apparatus with the rest of the muscle cells is desmin. This desmosomal protein (described in more detail the next section 1.3.1) provides a network which interconnects the Z-discs of neighbouring myofibrils as well as connecting the myofibrils to the intercalated discs, the extracellular matrix and to the organelles and nucleus of cardiomyocytes (Small et al., 1992).

Structural and functional connection between adjoining cardiomyocytes in the heart is primarily maintained by the intercalated disc. In addition to gap junctions, which allow electrophysiological coupling of cardiomyocytes, the intercalated disc also contains a network of cell adhesion complexes such as desmosomes and adherens junctions (Figure 1.5). These two structures function by interconnecting neighbouring cardiac muscle cells as well as acting as membrane anchorage sites for myofibrils *via* the cytoplasmic actin and desmin cytoskeleton.

Mutations in genes encoding protein components of either of these two complexes have been linked to cardiac defects in humans such as Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC) (Awad et al., 2008) and Dilated Cardiomyopathy (DCM) (Perriard et al., 2003). Both these conditions involve a weakening of the ventricles of the heart caused by varying degrees of ventricular wall thinning and dysmorphogenesis, as well as dilation of one or both ventricular chambers (particularly in dilated cardiomyopathy). These morphological changes result in a decrease in pumping efficiency and wall stress resistance in the heart.

In ARVC, the thinning of the right ventricular wall is also linked with fibro-fatty replacement of the myocardium in this heart chamber in addition to arrhythmia. Whilst ARVC has been predominantly linked with a defect in desmosome assembly, DCM is generally regarded as a condition caused by deregulation in the contractile actin cytoskeleton tethering to the intercalated disc and defects in adherens junction proteins (Awad et al., 2008; Perriard et al., 2003).

1.3.1 Desmosomes

Desmosomes are abundant in tissues which experience mechanical stress, such as heart and skin, where they predominantly function to provide structural stability by forming adhesive complexes which tether cells together. This function is facilitated by multiple proteins interacting to form a macromolecular complex, also called plaque, which links the intermediate filaments of one cell to the intermediate filaments of another through transmembrane proteins (Figure 1.6 A).

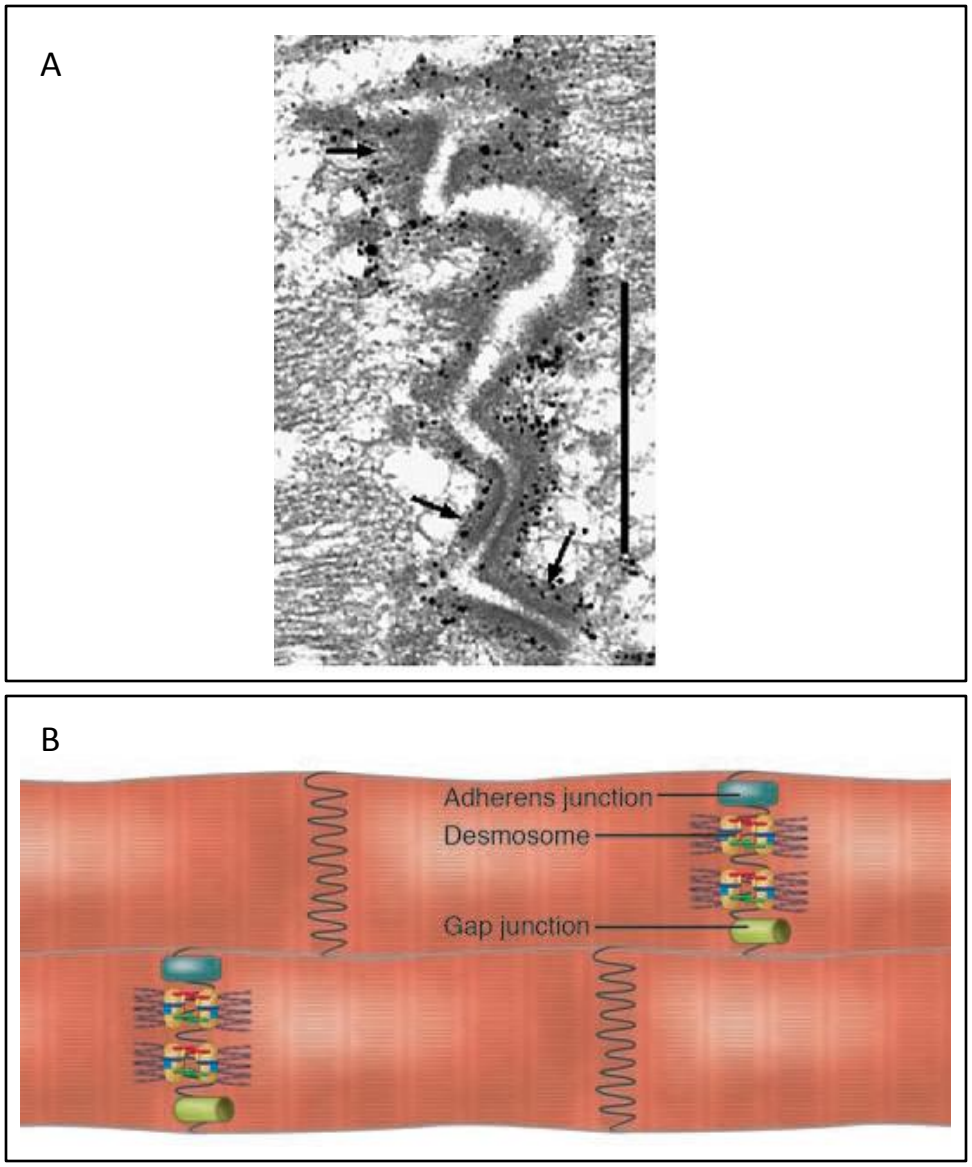


Figure 1.5: Structure of the intercalated disc in cardiomyocytes.
(A) Immuno-electronmicrograph of a cross-section through the intercalated discs of rat heart muscle showing labeling of the intercalated disc structure with desmoplakin antibodies. Desmosome complex regions (denoted by arrows) show a higher density of labeling antibody. (B) Schematic of the intercalated disc in cardiac muscle. The intercalated disc contains three types of membrane junction: gap junctions, desmosomes and adherens junctions. Bar in (A) = 500µm. Figure (A) was adapted from Franke *et al.* (2005) and figure (B) was adapted from MacRae *et al.* (2006).

The transmembrane part of the desmosome is made up of two distinct cadherin proteins - desmocollin and desmoglein. These proteins interact with the intracellular linker protein plakoglobin (also known as γ-catenin), which in turn is bound by the N-terminal region of desmoplakin. The desmosome plaque is connected to the intracellular intermediate filaments (desmin in cardiac cells) *via* the carboxy-terminal end of desmoplakin. The stability of this interaction is enhanced by the simultaneous binding of plakophilin to plakoglobin (Green and Gaudry, 2000).

As previously highlighted, the loss of structural integrity caused by mutations in desmosomal proteins is linked to human disease. Mutations in both *plakoglobin* and *desmoplakin* have been shown in Naxos disease, which causes woolly hair, skin blistering (keratoderma) and ARVC (McKoy et al., 2000; Norgett et al., 2000; Protonotarios et al., 2001). Less severe defects, such as isolated ARVC without hair and skin involvement, have been observed in human mutations of both *desmoglein-2* (Awad et al., 2006; Pilichou et al., 2006), *plakophilin-2* (Gerull et al., 2004) and *desmocollin* (Heuser et al., 2006). Currently only 50% of all cases of ARVC in humans have been linked to mutations in known desmosomal proteins, whilst among the identified desmosome components, mutations in *Plakophilin-2* appear to be the most prevalent cause of ARVC (van Tintelen et al., 2006).

The generation of loss-of-function alleles for desmosome components in the mouse model (Figure 1.6 B (Table)), and gene silencing in zebrafish have facilitated a better understanding of the functional requirement of desmosomes in the heart and the subcellular structural defects caused by the loss of plaque stability.

Deletion of *plakoglobin* (Bierkamp et al., 1996; Ruiz et al., 1996) and *plakophillin-2* (Grossmann et al., 2004) in mouse has revealed that both protein products are essential for the maintenance of structural integrity in the heart. Absence of either gene leads to a thinning of the compact myocardium and a lack of trabeculation in homozygote mutant animals, which can result in rupturing of the right ventricular heart wall and embryonic death.

Similar results have been observed in cardiac-specific deletion of *desmoplakin* (Garcia-Gras et al., 2006). Desmoplakin has been shown to be required for the structural integrity of the early blastocyst stage embryo (Gallicano et al., 1998), as well as the heart and epithelia of the skin and neural system in later stages

of development (Gallicano et al., 2001). Desmoplakin deficiency in the hearts of mutant animals resulted in death within two months from birth, caused by right ventricular dilation and contractile dysfunction due to fibro-fatty replacement (Garcia-Gras et al., 2006).

Mice heterozygous for a mutation in either *plakoglobin* or *desmoplakin*, although normal at birth, were also found to develop right ventricular dilation and dysfunction, leading to death at six months of age due to arrhythmia. This condition was exacerbated by exercise (Garcia-Gras et al., 2006; Kirchhof et al., 2006).

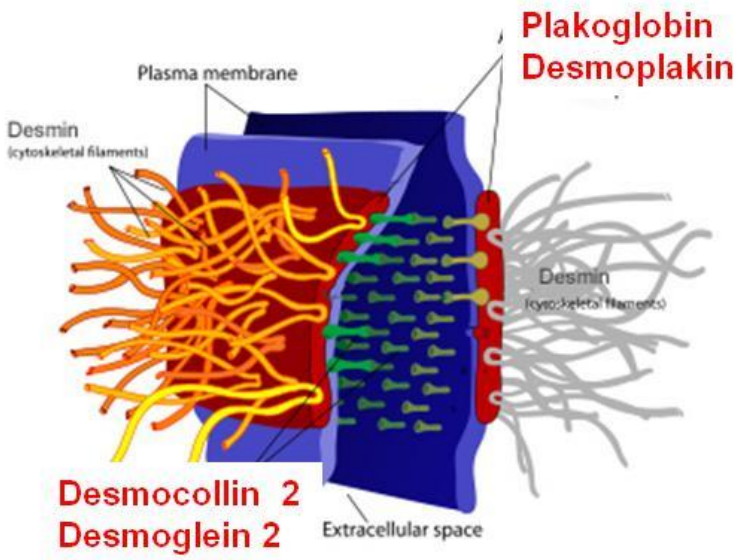
Mouse models of mutations in desmoglein or desmocollin have not been reported in the literature. However, silencing of desmocollin-2 by morpholino knockdown in the zebrafish model (Heuser et al., 2006) indicates that this protein may also be important for desmosome function, as zebrafish mutants display ventricular chamber dilation and contractile defects.

Detailed ultrastructural analysis of the intercalated disc in the mutant animal models has revealed that the loss of desmosome complex proteins, such as plakophilin-2 and plakoglobin, results in an uncoupling of the remaining plaque structure from the plasma membrane (Bornslaeger et al., 1996; Ruiz et al., 1996). Most strikingly, in *plakophillin-2*$^{-/-}$ animals, desmoplakin was found to dissociate from the intercalated disc and accumulate in the cytoplasm (Grossmann et al., 2004). Loss of the desmosome plaque in the heart results in a decrease of cell-cell adhesion at the intercalated disc, leading to impaired myocardial tissue integrity under conditions of mechanical stress.

In addition to the loss of heart tissue integrity, plakoglobin, which dissociates from the desmosome plaque in mutant animals, has been proposed to cause ARVC by competing with β-catenin in the Wnt pathway to alter the fate of cardiomyocytes to that of adipocytes (Garcia-Gras et al., 2006; MacRae et al., 2006).

As described in this section, mutations in desmosomal proteins are generally associated with ARVC in both humans and mice. However, this is not the case with mutations in the desmosome-associated intermediate filament: desmin. The loss of desmin in homozygote knockout mice results in dilated cardiomyopathy (Li et al., 1996; Milner et al., 1996) comparable to DCM observed in humans (Milner et al., 1999; Thornell et al., 1997). Although *desmin*$^{-/-}$ mice appeared to be normal at birth,

A



B

Mutations in Desmosome complex components in the mouse model

Gene	Type of mutation	Heart defect	Other defects	Reference
Plakoglobin	t, null	- death at E10.5 - RVW thinning & rupture - impaired contractility - absence of desmosomes	- skin blistering	(Bierkamp et.al, 1996) (Ruiz et.al, 1996)
Desmoplakin	t, null ES + Extraembryonic rescue	- death at E10 - heart malformation and collapse	- skin blistering - neural defects	(Gallicano et.al, 2001)
	t, null, cardiac muscle specific	- death at E10-12 some survival until 2 weeks post-partum - hypoplastic ventricles - adipocyte infiltration		(Garcia-Gras et.al, 2006)
Plakophilin-2	t, null	- death at E11 - RVW thinning & rupture - reduced trabeculation - disarrayed cytoskeleton		(Gross-mann et.al, 2004)
Desmin	t, null	- late onset dilated cardiomyopathy - myocyte cell death - exercise induced heart failure or even rupture	- muscle fatigue - perturbed myofibrill alignment in myocytes	(Li et al, 1996) (Milner et. al, 1996)

t – targeted mutation RVW – right ventricular wall

Figure 1.6: Protein components of the desmosome complex. (A)Schematic of the desmosome complex and its protein components tethering two adjacent cells. (B) Table of mouse phenotypes resulting from mutations in desmosome complex components. Schematic in (A) adapted from the public gallery of Ruiz, M. (2008) (http://neuroendokrinologie.de/portfolio6.htm)

defects in highly exercised muscle, such as the heart and diaphragm, start developing as early as 2 weeks of age (Li et al., 1997). Exercise in these animals can result in increased muscle fatigue. The hearts of these animals develop cardiac chamber dilation and myocyte cell death, leading to cardiac failure and in extreme cases exercise-induced ventricle wall rupture (Li et al., 1997).

All three muscle types - smooth, cardiac and skeletal - are affected by the loss of desmin (Li et al., 1996; Thornell et al., 1997). Structural analysis of desmin null myocytes showed that the alignment of the myofibrils in these cells was perturbed and neighbouring muscle cells had lost adhesion as seen by increased intercellular spaces. The reason for the development of DCM instead of ARVC in desmin null mice most likely lies in the fact that desmin is not only involved in the desmosome-dependent tethering myocyte cells, but also in stabilizing the sarcomeric cytoskeleton by tethering it *via* its Z-disc structures to the organelles and cell membrane of the muscle cell (Li et al., 1997). Thus the desmin network is important for contractile force transmission within cardiomyocytes and throughout the heart. Loss of maintenance of the contractile apparatus and inadequate force transmission, within and amongst cardiomyocytes, is generally linked to dilated cardiomyopathy (Capetanaki, 2000).

1.3.2 Adherens Junctions

In addition to desmosomes, a second membrane spanning protein complex, namely the adherens junction, has been shown to be required for providing adequate cell-cell adhesion between cardiomyocytes at the intercalated disc.

Adherens junctions function by tethering actin microfilaments to the plasma membrane (Figure 1.7 A). This function is achieved by a complex containing N-cadherin, which spans the intercellular space and is coupled *via* its cytoplasmic tail to both plakoglobin and β-catenin. These two proteins in turn interact with α-catenin, which binds to actin filaments anchoring them at the plasma membrane (Perriard et al., 2003). The anchoring of actin filaments *via* α-catenin is thought to be facilitated by a number of mediator proteins such as α-actinin (Blanchard et al., 1989; Djinovic-Carugo et al., 1999) and two members of the LIM-domain protein family: α-actinin

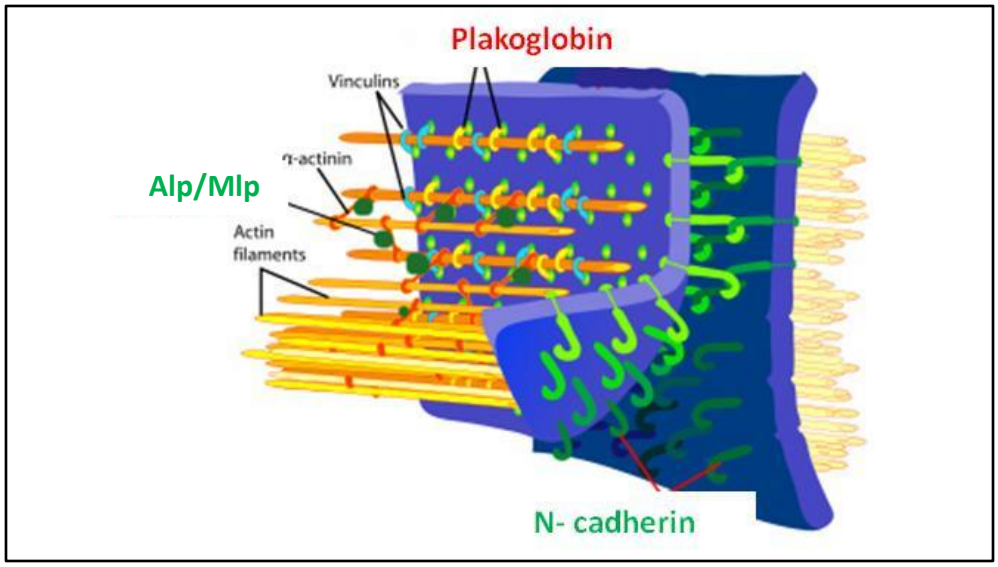
associated LIM-domain protein (Alp) (Pashmforoush et al., 2001) and muscle specific LIM-domain protein (Mlp) (Arber et al., 1994).

Adherens junctions are the first cell-cell adhesion complexes established during development and have been shown to be required to initiate the assembly of desmosomes (Huen et al., 2002; Linask, 1992). In fact, recent structural analysis of the intercalated disc indicates that both cell adhesion complexes interact closely and may even form mixed complexes at the intercalated disc (Borrmann et al., 2006; Franke et al., 2006). That plakoglobin has been found to be a major component in both structures (Ruiz et al., 1996) appears to support these findings as does the co-localisation of plakophillin-2 and desmoplakin with adherens junctions in the rat myocardium shown by these authors. Gene mutagenesis studies of adherens junction proteins in the mouse model have confirmed the importance of this complex for the structural integrity of the developing heart.

N-cadherin is the only cadherin expressed in cardiac muscle. Deletion of this protein specifically in the mouse heart at six to ten weeks after birth was shown to lead to the disassembly of the intercalated disc, including desmosome and adherens junction structures resulting in the subsequent death of the homozygote animals after the second month (Kostetskii et al., 2005).

Similar to N-cadherin, deletion of the actin cross-linking mediators *Mlp* and *Alp* have been found to affect the structure and function of the cardiac muscle in the ventricles (Ehler et al., 2001; Knoll et al., 2002; Pashmforoush et al., 2001). In particular, the phenotype in *Alp* deficient mice appears to be reminiscent of that seen in *Mospd3* gene trap animals (described in section 1.1.3).

Both Alp and Mlp are members of the PDZ-LIM-domain proteins, a group of proteins involved in the organization of the actin cytoskeleton (Zheng and Zhao, 2007). Alp and Mlp were both found to localize to the Z-discs of sarcomere structures in myocytes as well as being highly concentrated at the intercalated disc in cardiac muscle (Arber et al., 1997; Pomies et al., 1999; Xia et al., 1997). This localisation coincides with that of α-actinin. Biochemical studies particularly with Alp demonstrate that Alp and Mlp act by enhancing α-actinin mediated cross-linking of actin filaments and actin anchorage to the adherens junctions of the intercalated disc (Pashmforoush et al., 2001).



B

Mutations in Adherens Junction components in the mouse model

Gene	Type of mutation	Heart defect	Other defects	Reference
Alp	t, null	- neonatal death in 15% of animals - RVW thinning - RV dilation - decreased trabeculation		(Pashm-foroush et.al, 2001)
Mlp	t, null	- some death at 2 weeks - in surviving adults: - dilated cardiomyopathy - LV dilation - impaired contractile function	- fatigue	(Arber et.al , 1997)
N-cadherin	t, null	- death at E10 - severe cardiac malformation	- mild defects in somitogenesis & neuronal development	(Radice et.al, 1997)
	t, null, induced cadiac -specific deletion at 6-10 weeks	- death after 2 month - dilated cardiomyopathy - disarrayed cytoskeleton		(Kostetskii et.al, 2005)

t – targeted mutation LV- left ventricle RV – right ventricle RVW – right ventricular wall

Figure 1.7: Protein components of the adherens junction complex. (A)Schematic of the adherens junction complex and its protein components tethering two adjacent cells. (B) Table of mouse phenotypes resulting from mutations in adherens junction complex components. Abbreviations: Alp, α-actinin associated LIM-domain protein; muscle specific LIM-domain protein. Schematic in (A) adapted from the public gallery of Ruiz, M. (2008) (http://neuroendokrinologie.de/portfolio6.htm).

Mlp is expressed during differentiation of all striated muscle types, but expression is only maintained in cardiac and a subset of skeletal muscle in adult mice (Arber et al., 1994; Schneider et al., 1999). Loss of Mlp in the knockout mouse heart results in the loss of ordered sarcomeric arrangement along with morphological defects in cell-cell and cell-matrix connections (Arber et al., 1997). In order to compensate for the lack of Mlp at the ICD, the expression of other adherens junction proteins was found to be upregulated, presumably in an attempt to compensate for the loss of actin anchorage and the resulting decrease in force transmission in cardiomyocytes (Ehler et al., 2001).

The dilated cardiomyopathy observed in *Alp*$^{-/-}$ mice is less severe than that seen with the loss of Mlp. Instead of affecting both ventricles, the cardiomyopathy in *Alp*$^{-/-}$ mice appeared to be exclusive to the right ventricle. This chamber specificity correlates with the expression pattern of Alp. Whilst the protein is expressed throughout the heart at the early stages of cardiac development, it becomes restricted predominantly to the right ventricle wall and outflow tract after heart looping (Pashmforoush et al., 2001).

Lack of Alp in the mouse model leads to right ventricle wall thinning, RV chamber dilation and a decrease in trabeculation similar to the defect seen in *Mospd3*$^{Gt1lfm/Gt1lfm}$ animals (*Mospd3*$^{Gt1lfm/Gt1lfm}$ phenotype described in section 1.1.4). In both *Alp*$^{-/-}$ and *Mospd3*$^{Gt1lfm/Gt1lfm}$ mutants, only a percentage of homozygote neonates die, 15% and 50%, respectively (Pall et al., 2004; Pashmforoush et al., 2001). This neonatal mortality is in both cases dependent on the genetic background of the mouse strain carrying the genetic alteration. The presence of either mutation on a pure C57BL/6 background results in neonatal death whilst homozygous mutants on a mixed genetic background, such as MF1, are unaffected.

Alp has been shown to be important for α-actinin-associated actin cytoskeleton organization in cardiomyocytes (Pashmforoush et al., 2001). By facilitating actin fibre bundling through α-actinin, Alp helps to enhance the stability and mechanical resistance to high hemodynamic forces specifically in the embryonic right ventricular heart chamber (Lorenzen-Schmidt et al., 2005). Overexpression of Alp in cultured cardiomyocytes was found to improve sarcomeric structure and Z-disc arrangement (Pashmforoush et al., 2001).

Recent reports also show that, in addition to stabilizing the sarcomeric complex and thus providing efficient structural integrity and contractile force transmission in the heart, Alp is also required for myocyte differentiation *in vitro*. The α-actinin-associated actin filament stabilization through Alp was found to provide an important structural signal for MyoD (myogenic differentiation) and myogenin-dependent differentiation of myocytes in culture (Pomies et al., 2007).

With the exception of α-actinin and Mlp (Mohapatra et al., 2003), mutant forms of other adherens junction proteins have not currently been linked to cardiomyopathies in humans.

1.4 FUNCTIONAL ANALYSIS OF GENES

1.4.1 Using animal model systems

Animal models have played an important role in the study of genes and diverse biological aspects. Model organisms including worms, flies, fish, frogs and rodents have proved to be invaluable tools for understanding how genes work, and for elucidating fundamental principles of development. Among these organisms, the mouse has become a favourite model for researching mammalian biology such as embryonic development, disease, behaviour and cancer. Several features have helped its popularity as the preferred mammalian model organism. Short generation cycles, relatively large litter sizes and a small body make it a very cost effective model. More importantly however, mice show a high degree of physiological similarity to humans which is reflected in the similarity between the mouse and human genome. Analysis of these two fully sequenced genomes indicates that almost all genes in the human have mouse counterparts and *vice versa*.

A similar level of genomic conservation to human is also seen in rats, which have long been the preferred organism for modelling human physiology. But there are two unique feature of the mouse model setting it apart from other mammalian model systems. These are the availability of sophisticated genetic manipulation technologies, and well established and characterized embryonic stem (ES) cell lines.

1.4.2 ES cells

The establishment of mouse embryonic stem cells has opened many exciting experimental approaches in the field of mammalian developmental biology.

ES cells are pluripotent cells derived from the inner cell mass (ICM) of the blastocyst-stage mouse embryo (Evans and Kaufman, 1981; Martin, 1981). These cells can be maintained in feeder-free culture indefinitely due to their ability to self renew when grown in the presence of serum and leukaemia inhibitory factor (LIF) (Smith et al., 1988). When re-introduced into a host blastocyst, ES cells also have the ability to differentiate into all cells of the embryo proper (Keller, 1995), allowing mutations introduced into ES cells to be carried through to the germ line, in order to facilitate the generation of mice with specific genetic modifications.

In addition to their differentiation capabilities *in vivo*, ES cells can be induced to differentiate into a wide range of cell types *in vitro*. The differentiation of ES cells in culture can facilitate the isolation and study of early differentiation precursor cells, as well as help to assess the differentiation potential of ES cells carrying mutations in genes essential for embryonic development.

The differentiation of ES cells *in vitro* can be achieved by generation of embryoid bodies (EBs) (Doetschman et al., 1985; Keller, 1995). These structures are formed by aggregation of ES cells in a 'Hanging Drop' culture. Removal of LIF from EBs in culture results in spontaneous differentiation of these structures, promoting the formation of a range of differentiated cell types such as hematopoietic cell lineages (Doetschman et al., 1985; Nakano et al., 1994; Nishikawa et al., 1998), endothelial cells (Risau et al., 1988; Yamashita et al., 2000), neuronal cells (Fraichard et al., 1995; Strubing et al., 1995) and bone derivatives (Buttery et al., 2001; Kramer et al., 2000).

1.4.3 *In vitro* Cardiomyocyte Differentiation Model

When differentiating EBs are allowed to attach to gelatine they can often be observed to form foci of cells that begin rhythmic contraction, indicating cardiac muscle development.

The development of these cardiomyocyte lineages has been shown to progress through distinct stages, recapitulating the *in vivo* process of differentiation. Akin to *in vivo* development, beating cardiomyocytes in culture express cardiac gene products in a developmentally controlled manner, initially expressing lineage induction markers such as *Gata4* and *Nkx2.5* followed by distinct cardiomyocyte maturation genes such as *Nppa* (natriuretic peptide precursor type A), *Mlc2v* (myosin light chain 2v), myosin heavy chains and connexin 43 (Boheler et al., 2002; Hescheler et al., 1997). As in the *in vivo* system, the differentiation process in cardiomyocytes is associated with morphological cell shape changes from round cells towards elongated cells with well developed myofibrils and sarcomere structures (Robbins et al., 1990; Westfall et al., 1997). Furthermore, the shift from early pacemaker cells to more terminally differentiated atrial- and ventricle-like cell

types indicates that differentiating cardiomyocytes in culture can correctly mirror the electrophysiological changes observed *in vivo* (Hescheler et al., 1997; Maltsev et al., 1993).

The ability to induce multiple cell lineages during *in vitro* differentiation of ES cells provides exciting opportunities to, not only model embryonic development in culture, but also study the developmental potential of deliberate mutation in genes implicated in lineage specific development.

Accordingly, the *in vitro* cardiomyocyte model has been successfully used to elucidate the role of transcription factors and ECM components in cardiac development. Loss-of-function studies of Gata4 and Hand1 transcription factors by *in vitro* differentiation of ES cells into cardiomyocytes have revealed that, although both genes appear to be crucial for heart morphogenesis as indicated by the death of *Gata4* $^{-/-}$ and *Hand1* $^{-/-}$ embryos *in utero* from myocardial defects (Firulli et al., 1998; Kuo et al., 1997; Molkentin et al., 1997), neither of these genes is required in order to generate beating cardiomyocytes in culture (Narita et al., 1997; Riley et al., 2000). This result is supported by the finding that *Gata4* $^{-/-}$ and *Hand1* $^{-/-}$ ES cells can contribute to the developing heart in chimaeric embryos (Narita et al., 1997; Riley et al., 2000).

Proper myocardial development does not only depend on transcription factors but also on the interaction of ECM components with the cytoskeleton of the developing cardiomyocytes. Whilst the loss of the cardiac-specific intermediate filament - desmin - has barely any effect on the *in vitro* cardiomyocyte commitment and differentiation potential (Weitzer et al., 1995), loss of β1-integrin (providing cardiomyocyte and ECM contact) leads to delayed cardiomyocyte differentiation in conjunction with disarrayed sarcomere structures and reduced survival of terminally differentiated myocardial cells (Fassler et al., 1996). Hence, β1-integrin appears to be required for terminal differentiation and maintenance of cardiac muscle cells (Fassler et al., 1996; Guan et al., 2001).

1.4.4 Gene Mutagenesis in ES cells

The generation of mutations in the mouse genome is arguably one of the most powerful tools for studying gene function. The scope of such genetic manipulations has been greatly expanded since the isolation of ES cells. For instance, the availability of ES cells has been instrumental to the development of mouse genome mutagenesis approaches such as gene trapping (Gossler et al., 1989) and gene targeting (Thomas and Capecchi, 1987).

The mutagenesis of genes in ES cells has several advantages. Firstly, ES cells can be screened *in vitro* for the desired genetic alteration before being reintroduced into the embryo. Secondly, it is possible to study the effect of genetic manipulation on the differentiation potential of ES cells using the previously described *in vitro* differentiation models (described in section 1.4.2 and 1.4.3).

1.4.5 Gene trapping

Gene trapping in ES cells is one of the most widely used mutagenesis approaches. This phenotype-driven method enables large-scale mutagenesis in mice combined with relatively easy cloning and identification of the mutated gene. In its most basic form the vector used for gene trapping contains a splice acceptor site immediately upstream of a promoterless reporter gene such as *β-galactosidase* (*β-gal*), which is followed by a selectable marker gene as well as a polyA tail for transcriptional termination (Figure 1.8) (Gossler et al., 1989).

The integration of this vector into an active gene can lead to the expression of an in-frame fusion transcript between the upstream sequences of the gene and the *β-gal* reporter gene (Figure 1.8) (Friedrich and Soriano, 1991; von Melchner et al., 1992). The presence of a transcription termination site in the trapping vector prevents the downstream sequences of the gene from being expressed, thus generating a truncated mutant transcript. In addition, the presence of the *β-gal* reporter under the control of the trapped genes promoter allows analysis of the genes endogenous expression pattern whilst providing a convenient DNA tag for the identification of the trapped locus (Stanford et al., 2001).

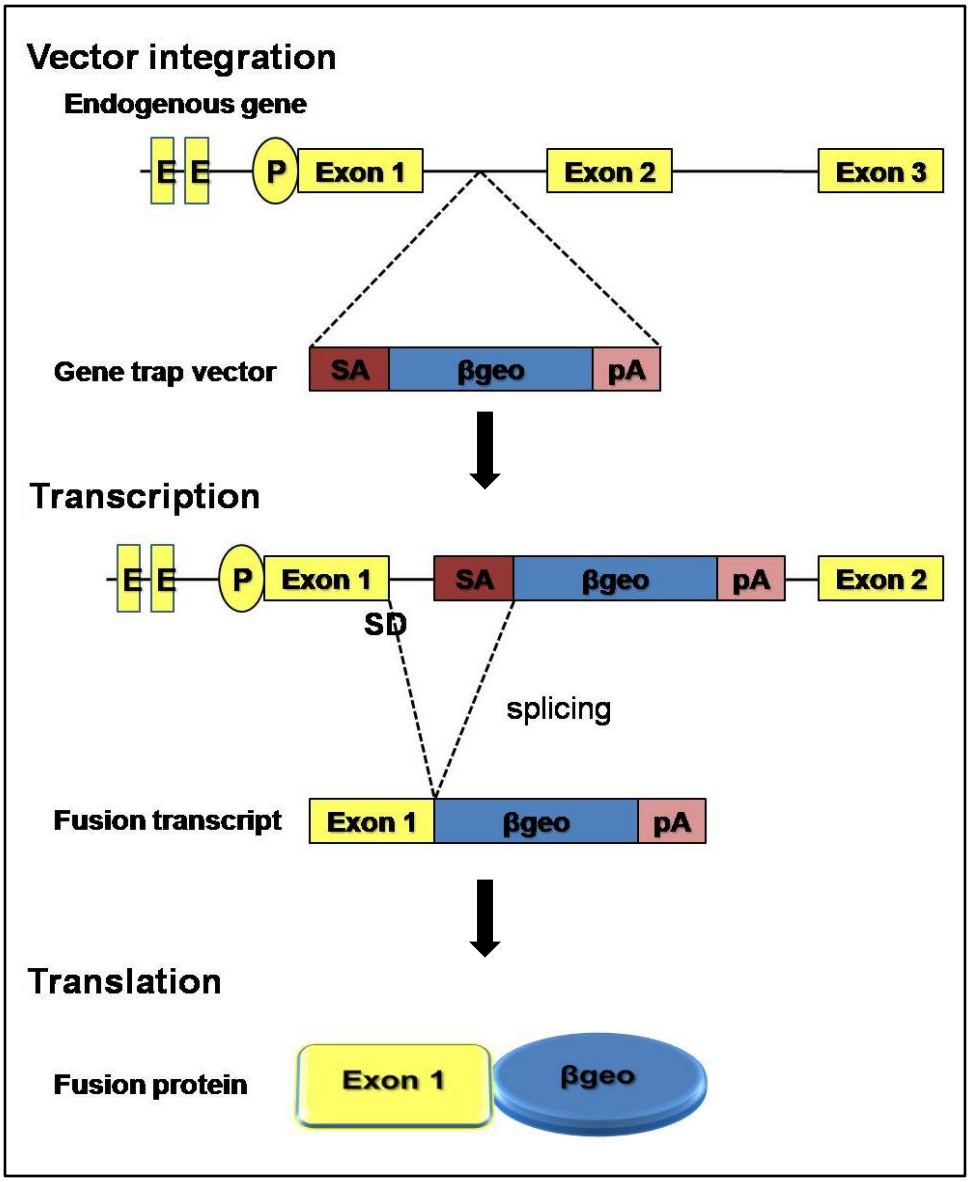


Figure 1.8: Gene trapping. The gene trap vector containing a splice acceptor (SA), *βgeo* (*β-galactosidase* reporter gene fused to a neomycin resistance gene) and a polyadenylation signal (pA). Random integration of this vector into the intron of a gene results into the generation of a fusion transcript between the vector and the trapped endogenous gene. The splice donor of the upstream exon (exon1) is spliced to the splice acceptor sequence present in the vector transcript. Transcription of the vector sequences is regulated by enhancer and promoter elements of the trapped endogenous gene. Translation of this fusion transcript results in a fusion protein. Abbreviations: E, enhancer; P, promoter; pA, polyadenylation signal; SA, splice acceptor; SD, splice donor.

A number of centres world-wide have performed high-throughput gene trapping screens. ES cell clones generated in these screens have been made publicly available through the International Gene Trap Consortium (IGTC)[4]. This repository currently comprises nearly 135,500 ES cell clones, representing more than 32% coverage of the mouse genome. The International Gene Trap Consortium database allows researchers to search for suitable gene trap integrations into their gene of interest before generating a gene mutation of their own.

A search of the database for gene trap integrations into *Mospd* gene found 5 different vector integrations into *Mospd3*, and none in *Mospd1*, *Mospd2* or *Mospd4*. One of these gene trap integrations appears to correspond to the vector integration into the 5'-untranslated gene sequence, also observed in the *Mospd3*^{Gt1lmf} allele. This integration led to the identification of *Mospd3* (described in section 1.1.3). In the remaining four ES cell lines the vector appears to have integrated into the last or penultimate exon of the gene, leading to a likely loss of one, or both, of the transmembrane domains of Mospd3.

The most effective gene trap integrations usually involve vector insertions near the 5' end of a gene, resulting in early termination of the endogenous gene transcription. On the other hand, integrations into 3' introns can sometimes result in incomplete inactivation of the protein, since upstream sequences are still expressed; thus, leading to reduced (hypomorphic) function (Herrick and Cooper, 2002; McClive et al., 1998; Skarnes et al., 1992). As gene trap integrations occur randomly in the genome, a large number of mutations can be generated. To date, it is estimated that about two-thirds of all known protein coding genes in the mouse have been trapped using this approach (Skarnes et al., 2004).

1.4.5.1 Gene trap screens for developmentally regulated genes

The use of gene trap mutagenesis in the mouse facilitates the identification of developmental genes expressed in a specific cell lineage. However, this *in vivo* approach requires the generation of large numbers of mice from trapped ES cell clones, to identify only a limited number of developmentally regulated genes.

[4] www.genetrap.org

Fortunately, the ability to differentiate ES cell *in vitro* into a wide range of different cell lineages (as previously described in Section 1.4.2) has helped to overcome this problem, and has enabled high-throughput screening of ES cell clones for tissue-restricted gene trap expression. Using this *in vitro* screening approach, genes expressed specifically in neuronal, hematopoietic and endothelial lineages; as well as in myocytes and chondrocytes; have been successfully identified (Baker et al., 1997; Shirai et al., 1996; Stanford et al., 1998).

1.4.5.2 Inductive gene trap screens

An alternative approach to identifying developmentally regulated genes, *via* gene trap reporter expression in *in vitro* differentiated ES cells, is to induce differentiation of ES cells into defined cell lineage *in vitro*, by provision of specific developmental cues in the form of growth and differentiation factors.

One such inductive gene trap screen was designed to isolate genes responsive to transforming growth factors such as BMP2, nodal and activin. This enabled identification of *Chondroitin-4-sulfotransferase*. This gene, which was induced by BMP2, was subsequently shown to exhibit a distinct spatio-temporal expression pattern in embryonic development (Kluppel et al., 2002). Enrichment for developmentally regulated genes was also achieved in another induction screen, in which gene trapped ES cell clones were analysed for their response to follistatin, nerve growth factor and retinoic acid (Bonaldo et al., 1998).

1.4.5.3 Induction trap used to identify *Mospd3*

The high-throughput induction gene trap screen which led to the identification of *Mospd3* was designed to isolate genes which lie downstream of the retinoic acid (RA) receptor ligand-mediated signalling pathway (Forrester et al., 1996). Retinoic acid, the active metabolite of vitamin A, has been shown to affect the expression of developmental genes in embryos (Conlon and Rossant, 1992; Marshall et al., 1992), as well as ES cells differentiated in culture (Wobus et al., 1997).

In this screen, trapped ES cell clones were selected which exhibited an RA induced change in *β-gal* expression upon differentiation in culture. In total 20 trapped genes were identified in this study. Among these clones, nine showed an

induction of *β-gal* activity whilst the remaining 11 exhibited repression in response to externally provided RA. Furthermore, ten of the repressed clones also showed a spatially restricted pattern of reporter gene expression in the developing embryo *in vivo*. The *β-gal* expression pattern, observed in repressed clones, appeared to be distinct from that of induced clones. One of these ten repressed gene trap integrations was shown to have trapped a novel gene, called *Mospd3*. Like most of the repressed lines, the *Mospd3* gene trap reporter showed a distinct expression in the embryonic heart (Forrester et al., 1996; McClive et al., 1998).

1.4.5.4 Complex gene trap integrations

Gene trap mutagenesis is a very powerful tool for the identification of new genes, and their functional analysis. However, not all gene trap integrations are as straightforward as previously described. Some of these unsuspected complexities have been analysed and reviewed (Voss et al., 1998). In one case, Niwa *et.al.* found that the gene trap vector, used in their screen, caused large deletions and rearrangement in the DNA sequences surrounding the vector integration site (Niwa et al., 1993). In another screen, the gene trap vector was found to trans-splice to splice donor sites unrelated to the trapped gene (Voss et al., 1998).

In the case of the gene trap screen which resulted in the identification of *Mospd3*, detailed analysis of the integration site revealed that the vector had integrated at the start of the first exon of *Mospd3* (Pall et al., 2004). This exon contains the 5'-untranslated region of the gene, including part of a highly conserved sequence (83% similarity between mouse and human). This conserved region, stretching from sequences upstream of *Mospd3* into exon1 of the gene, corresponds to the putative promoter of *Mospd3*. It was shown that the expression of the *β-gal* reporter did not recapitulate all of the expression domains of the endogenous gene (Forrester et al., 1996). Similar observations have been made in other gene trap mouse lines (Deng and Behringer, 1995; Skarnes et al., 1992; Voss et al., 1998). In addition to having trapped the promoter of *Mospd3*, the vector has integrated in multiple copies in reverse transcriptional direction to that of the gene itself. Similar tandem vector integrations have also been reported (Friedrich and Soriano, 1991). The presence of other genes in close proximity to the vector integration site gave rise

to the possibility that, in addition to disrupting *Mospd3*, the integration also alters the transcriptional regulation of adjacent genes. Expression of neighbouring genes was examined, including that of *pcolce*, which was found to be unaffected by the integration.

1.4.6 Gene targeting

This method for mutagenesis of the mouse genome is based on the site-specific introduction of an exogenous DNA sequence into chromosomal DNA *via* homologous recombination. Gene targeting facilitates a wide range of precise target gene modifications including gene deletion, addition of new genes, and the introduction of subtle gene mutations such as base pair changes in the target sequence. If the gene targeting has been performed in ES cells, clones featuring the correctly targeted allele can be used to generate germ line transmitting transgenic mice to study the phenotype arising from this vector integration (Joyner, 1999; Thomas and Capecchi, 1987).

There are two types of vector, which can be used for target gene modification. Depending on the type of desired mutation to be generated, an insertion or a replacement vector can be used.

Insertion vectors have been predominantly used for target gene inactivation through integration of a mutagenic sequence. This type of vector can also be used to generate subtle target gene mutations in conjunction with a "hit and run" or "in and out" targeting approach detailed in section 1.4.6.2. Linearization of the insertion vector within the region of homology to the target site results in the integration of the entire vector into the endogenous DNA sequence. The integration of the vector through a single homologous recombination cross-over event also leads to a partial duplication of the targeting homology.

The replacement targeting vector has found much wider application than the insertion vector, especially since the development of conditional targeting constructs, which are described in more detail in sections 1.4.6.3 -1.4.6.7. Replacement vectors generally consist of a 5'- and a 3'- targeting homology sequence, flanking a mutant foreign sequence such as a selectable marker gene and other modifications.

1.4.6.1 Factors affecting gene targeting frequency

The rate of targeting vector integration upon electroporation into ES cells is very low (10^{-3}). It is therefore important to include a selectable marker in the targeting vector sequence, in order to enrich for ES cells carrying the vector integration. The most commonly used positive selection cassette, for this purpose, is the neomycin (neo) phosphotransferase gene under the control of a constitutive pgk (phosphoglycerate kinase 1) promoter. Other positive selection cassettes include hygromycin B (hygro)-phosphotransferase and the puromycin (puro) gene.

Most vector integration events into genomic DNA take place via non-homologous recombination. Only about one in every 1000 vector integrations occurs by homologous recombination (Thomas et al., 1986). One of the ways to select against random vector integration events in ES cells lies in using a negative selection approach in conjunction with positive selection. This positive-negative selection approach is based on a replacement vector containing the positive selection marker within the flanking regions of homology and a negative selection cassette at one or both ends of the targeting homology (Mansour et al., 1988). The most commonly used negative selection cassettes are the diphtheria toxin A (DT or DT-A) gene (Yagi et al., 1990) and the herpes simplex virus thymidine kinase (HSV-tk) gene. Isolation of ES carrying a homologous targeted integration is achieved by initial positive selection through application of neomycin, puromycin or hygromycin; followed by selection against ES cell clones carrying the negative selection cassette. Homologous integration of the targeting vector into chromosomal DNA *via* cross-over events in the flanking homology sequences, between vector and target site, results in the loss of the negative selection cassette. In the case of random vector integration, on the other hand, the cassette generally integrates along with the rest of the vector. The presence of the negative selection cassette leads to ES cell death through the expression of the diphtheria toxin A gene; or in the case of the HSV-tk cassette, cell death through the conversion of the nucleoside analogues ganciclovir or FIAU (Fialuridine) in the selective media into toxic metabolites. Positive-negative selection was found to improve the retrieval of correctly targeted ES cell clones only by about 5- to 10-fold (Jeannotte et al., 1991; McMahon and Bradley, 1990). Due to

occasional loss of the negative selection cassette during non-homologous recombination, the enrichment achieved by positive-negative selection is limited.

A number of additional factors, besides the positive-negative selection strategy, are known to improve retrieval of correctly targeted ES cell clones. These parameters include the type of target locus, the length of the homology region used in the targeting construct, as well as the use of isogenic DNA (Frohman and Martin, 1989). The length of the targeting homology sequences is known to have a particularly profound impact on the frequency of homologous vector integrations. Even small increases in the length of these sequences, up to a total length of about 10kb, can dramatically increase the frequency of gene targeting (Deng and Capecchi, 1992; Hasty et al., 1991b). Likewise, the use of isogenic rather than non-isogenic DNA for the generation of the targeting vector has been shown to increase the potential targeting frequency of a vector by 10- to 20-fold (te Riele et al., 1992).

1.4.6.2 Integration of subtle mutations

The ability to generate subtle mutations, such as single base-pair changes, in a target sequence is very important for modelling human genetic defects, which are often caused by single base-pair mutations in a gene.

The nature of this type of genetic modification required the development of a targeting strategy which resulted in a subtle mutation without leaving a selectable marker in the modified gene. Two different types of approach have been developed for this kind of gene targeting.

The first type of method called “hit and run” (Hasty et al., 1991a) or “in-out” (Valancius and Smithies, 1991) was based on the use of an insertion-type gene targeting vector. One of the characteristics of this type of vector is that it has the ability to self-excise from the locus it has been targeted to. This excision occurs via homologous recombination of the duplicated homology sequences which have been generated during the integration of the insertion vector into the target locus. Although it only occurs at very low frequency, this excision event can result in the removal of the endogenous gene sequences whilst leaving behind the modified gene targeting construct with the subtle mutation of choice. The alternative outcome in this case is that the targeted locus reverts to the endogenous allele, as the insertion vector containing the subtle mutation is excised instead.

The second type of approach, called “tag and exchange” (Askew et al., 1993) or “plug and socket” (Detloff et al., 1994) is suitable for generating subtle mutations using gene replacement-type vectors. These two-step gene targeting procedures involve the initial integration of a vector construct carrying both a positive and a negative selection cassette. In a second targeting step, the vector sequences carrying the selectable markers can be replaced by homologous recombination of a targeting construct containing the endogenous gene sequence with a single base-pair or other subtle mutation. In the “tag and exchange” approach, cells which have not been correctly targeted in the second vector integration step die if grown in negative selection medium. Negative selection thus enables the enrichment for correctly targeted clones.

The “plug and socket” approach contains a modification on the “tag and exchange” approach. Instead of using HSV-tk as the negative selection cassette, a non-functional mutated HPRT (hypoxanthine-guanine phosphoribosyltransferase) minigene is used in the first targeting construct, along with a positive selection marker. The second targeting construct, containing the mutated target gene sequence, also features a mutant HPRT minigene - in one of the flanking homology arms. Correct integration of the second targeting construct into the target locus leads to homologous recombination between the sequences of the two mutant HPRT minigenes, thereby creating a functional HPRT positive selection marker. The presence of a functional HPRT gene facilitates HAT-medium mediated enrichment of all ES cell clones which have integrated the mutated gene sequence. The HPRT minigenes of the two replacement vectors are designed in the flanking homology regions away from the gene so as not to interfere with the expression of the targeted gene.

Since the development of the Cre/loxP system (described in more detail in the following sections) it has become possible to generate subtle mutations in ES cells and subsequently delete the positive selection marker used for the enrichment of targeted clones.

1.4.6.3 Conditional control of gene expression

Constitutive inactivation of some genes may not be suitable for the functional analysis in the mouse model. If the deleted gene is crucial for embryonic development its loss can lead to embryonic lethality thus impeding the study of the genes function in later developmental stages (Mitchell et al., 2001) and adulthood.

It is therefore important to have the ability to use conditional systems which provide control over the spatial and temporal expression of a gene. Two main strategies have been developed for this purpose: 1) conditional gene knockout through site-specific recombination and 2) inducible expression of a transgene. Both methods provide control over when and where a gene is activated or silenced.

1.4.6.4 Site-specific DNA recombination

Site-specific recombination is currently the most widely used method for inactivation of a gene in a time and tissue-specific manner.

This system is based on two components: site-specific recombinase enzyme, and recombination recognition DNA sequences. Cre and Flpe are the most commonly used site-specific recombinases. Both enzymes belong to the integrase family of recombinases, and have been derived from bacteriophage P1 (Sauer and Henderson, 1989) and *Saccharomyces cerevisiae* (Buchholz et al., 1998; O'Gorman et al., 1991), respectively.

The function of these recombinases lies in catalysing the DNA recombination between two 34bp recognition sites such as *loxP* (acted on by Cre) and *FRT* (Flpe) (Nagy, 2000). These recognition sites consist of two 13bp inverse (palindromic) sequence repeats surrounding an 8bp core which assigns direction to the sequence (Figure 1.9 A). Depending on the orientation of two recognition sites - placed into a DNA sequence - site specific recombination can lead to deletion, inversion or switching of the sequence linking these sites. Both the Cre/loxP and the Flpe/FRT system are widely used; but, in further descriptions of the conditional allele, I will focus mainly on the former.

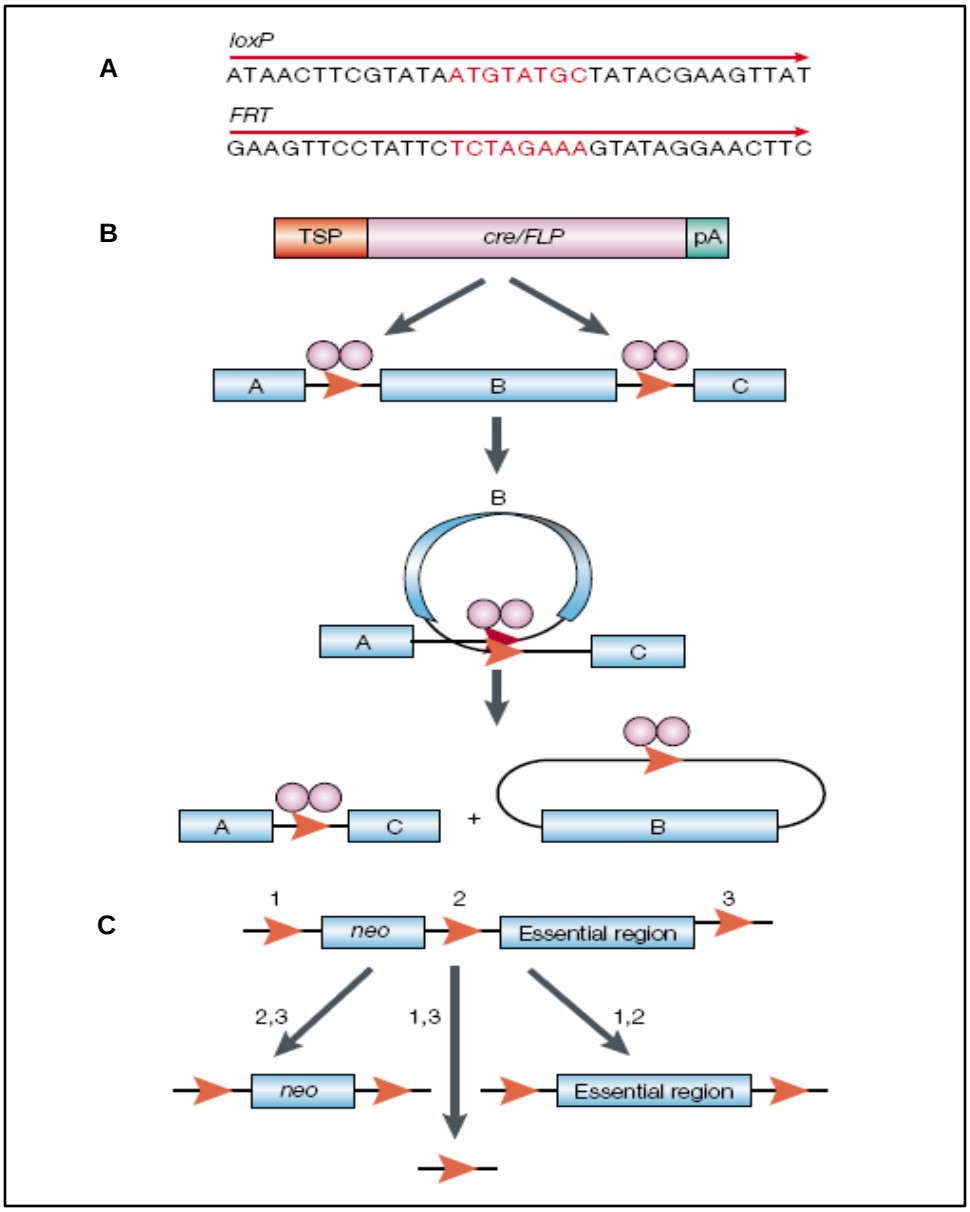


Figure 1.9: Site-specific recombination mediated by Cre or Flp recombinase. (A) The 34-bp *loxP* and *FRT* sites each consist of two 13-bp inverse repeats (black) that flank an 8-bp core sequence (red). This core sequence confers directionality to these sites (red arrows). (B) Dimers of Cre or Flp (pink) catalyze the recombination between two directly repeated *loxP* or *FRT* sites (red arrowheads), resulting in a deletion of sequences (region B) flanked by l*oxP* or *FRT*. If region B is an essential region of a gene, then the recombination event results in gene inactivation. (C) The use of three *loxP* sites can result in three different alleles following Cre or Flp recombination. Either the neomycin (neo) resistance cassette as well as the essential region of the gene are deleted (1, 3; generating a null allele) or only one of these regions is removed resulting in either a conditional f*lox*ed allele (1, 2) or a null allele of the gene leaving a neo cassette (2,3). Abbreviations: pA, polyadenylation site; TSP, tissue-specific promoter. Figure adapted from Lewandoski *et al*. (2001).

As the site-specific recombination of DNA sequences flanked by *loxP* sites (f*lox*ed) is dependent on the presence and action of Cre recombinase, it is possible to control this process by adding the enzyme to drive the reaction only when and where desired (Figure 1.9 B). This control is particularly advantageous in the generation of knockout alleles. Using this conditional system, early lethal phenotypes can be prevented by addition of Cre recombinase only at later stages of embryogenesis.

The design of a conditional knockout allele generally involves placement of the *loxP* sites upstream and downstream of an essential region of the gene of interest (Figure 1.9 C). Prior to Cre induced recombination, the gene should display wild type function. To avoid any effect of these *loxP* sites on the normal expression of the floxed gene, they are usually placed into the introns flanking the essential sequences of the gene.

1.4.6.5 Tissue specific control of *Cre* expression

The tissue and time-dependent inactivation of essential gene sequences in a conditional allele can be achieved by controlling the expression of Cre recombinase. Tissue specific inactivation of a gene of interest can be achieved by crossing the mouse line harbouring a conditional KO allele of the gene to an effector mouse strain which expresses *Cre* from a tissue specific promoter. In the resulting progeny, the gene of interest will only be inactivated in the tissues where Cre is expressed (Gu et al., 1994; Lewandoski, 2001). This conditional approach allows complex questions about the expression of a particular gene of interest to be asked.

Cardiac specific deletion of *Fgf8*, for example, has revealed that this growth factor is required for second (anterior) heart field development (SHF) (Ilagan et al., 2006). Fgf8 appears to be involved in four different processes of heart development. Firstly, Fgf8 is involved in the cardiac specification during gastrulation; secondly, in the establishment of right-left axis for cardiac looping; and thirdly, in promoting survival of neural crest cells migrating into the heart (Abu-Issa et al., 2002; Albertson and Yelick, 2005; Sun et al., 1999). Finally, Fgf8 has also been implicated in the specification and maintenance of the second heart field (Kelly et al., 2001). In order to show the requirement for Fgf8 in the SHF, the gene was inactivated in, both, the first and second heart field using Cre recombinase expressed from the cardiac specific Nkx2.5 promoter (Ilagan et al., 2006). Mice carrying this tissue specific

deletion of Fgf8 showed truncation of the right ventricle and the outflow tract. Deletion of Fgf8 specifically in the first heart field - and not the second heart field, using Cre expressed from the *titin* gene promoter, resulted in no apparent phenotype. The comparison of these two cardiac-specific knockouts has confirmed the role of Fgf8 in the SHF.

1.4.6.6 Inducible transgene activation

Several binary models have been developed to provide control over the transcription of a transgene in mice. These transactivation approaches are based on a transgene being placed under the control of a promoter/enhancer element which can be specifically regulated by a ligand-inducible transcription factor. The requirements for the transcriptional activation or silencing in such a system are; that the transactivation should be fast, robust; and should only occur if the effector molecule is present. The process should also be reversible by effector depletion. The most commonly used transactivation systems are the Tet "on/off" (Gossen et al., 1995) and the GAL4-based GLVP system (Ornitz et al., 1991).

The tetracycline system consists of two versions of the tetracycline controlled transactivator, which can either block (tTA/ tet "off") or activate (rtTA/ tet"on") the expression of a transgene in the presence of a doxycycline inducer. The presence of doxycycline in the tet"off" system prevents the tetracycline-controlled transactivator (tTA) from binding to the tetracycline response element which controls the expression of the transgene. Transgene expression is thus blocked in this setting. Binding of the tetracycline response element to reverse tTA (tet "on"), on the other hand, is promoted by the presence of doxycycline, thus allowing transgene expression to be activated.

The second inducible system, the GLVP system, is based on the transactivation of a target transgene by the Gal4 transcriptional transactivator from *Saccharomyces cerevisiae*. The Gal4 transactivator, in this system, has been engineered to contain a mutated steroid ligand domain. Due to this mutated steroid receptor component, Gal4 can only activate its target transgene in the presence of a suitable synthetic hormone such as the antiprogesterone compound RU486.

Similar transactivation systems can also be used to provide temporal control of Cre expression.

1.4.6.7 Temporal control of Cre expression

Temporal control of Cre mediate recombination can be achieved by combining the Cre/lox system with a transcriptional transactivation system such as Tet "on". Using this system, Cre expression can be activated in a defined manner by administration of doxycycline to mice (Kistner et al., 1996).

Another approach, providing temporal control over the activity of Cre recombinase, is the Cre-LBD (ligand-binding domain) system (Metzger et al., 1995). This post-translational transactivation system is based on a fusion construct of the Cre transgene with a mutated steroid ligand-binding domain under the control of a ubiquitous or tissue specific promoter. Upon expression, the Cre fusion protein is sequestered in the cytoplasm of the cell by a heat-shock protein. Binding by the heat-shock protein leaves the chimaeric Cre protein unable to localise to the nucleus, where Cre acts to mediate the recombination of the f*lox*ed target-gene allele. Nuclear localisation of Cre can only occur in the presence of a suitable synthetic steroid such as RU486 or tamoxifen (Kellendonk et al., 1996; Zhang et al., 1996). The ligand-binding domain in this system has been mutated to prevent endogenous steroid hormones, such as estrogen or progesterone, from translocating Cre to the nucleus.

One very popular version of this approach is the MerCreMer system (Sohal et al., 2001). The inducible *Cre* gene in this approach is flanked by two mutated murine estrogen-receptors. This system has been widely used in generating conditional gene deletions in adult mice. The inducible MerCreMer system has been successfully used in two different studies analysing the importance of adherens junction proteins in maintaining the structural integrity of the intercalated disc (Kostetskii et al., 2005; Zhou et al., 2007) (Adherens junctions have been described in detail in section 1.3.2).

The first of these studies involved the inactivation of β-catenin specifically in adult cardiomyocytes. B-catenin signalling controls a multitude of biological processes such as embryonic development, tumorigenesis as well as the establishment of cell shape and polarity (Klymkowsky, 2005). This wide range of activities can complicate the study of β-catenin function in a particular biological system. In order to circumvent the embryonic lethality caused by β-catenin inactivation and analyse the involvement of the protein in the adherens junction-

mediated structural integrity of the fully matured adult heart, an inducible conditional knockout approach was employed. Mice carrying a *loxP* flanked β-catenin allele were crossed with mice carrying a tamoxifen inducible MerCreMer transgene driven by a cardiac specific α-myosin heavy chain promoter. Using this inducible system, β-catenin was deleted only in the cardiac tissue of 3 to 4 month old mice after injection with tamoxifen. This study found that upregulation of γ-catenin can compensate for the absence of β-catenin from the adherens junctions in adult cardiomyocytes (Zhou et al., 2007).

Another study showed that cardiac-specific deletion of N-cadherin by MHC-MerCreMer, after tamoxifen administration to 2-3 month old mice, caused extensive remodelling of the intercalated disc; resulting in reduced cell-cell adhesion and contraction abnormalities in the hearts of these animals (Kostetskii et al., 2005).

The availability of tissue-specific promoters and inducible systems of Cre expression, to inactivate conditionally targeted genes, has made it possible to perform very intricate functional gene studies. The advent of these methods has also significantly widened the spectrum of possible modifications in the mouse genome. This flexibility was further facilitated by the discovery of novel recombinase enzymes (Diaz et al., 2001; Sauer and McDermott, 2004; Thyagarajan et al., 2001), and the development of mutant site-specific recognition sites which do not recombine with wild type ones (Araki et al., 2002).

1.4.6.8 Neomycin resistance marker and hypomorphs

The conditional targeting vector, usually, contains a selectable marker to allow selection of ES cells which have integrated the vector construct into their genome. Although it is a necessary part of the conditional construct, the presence of the *neo* resistance marker - in introns or in the untranslated sequences of a gene - can interfere with the expression of the f*lox*ed allele, creating a hypomorphic allele which shows reduced gene activity (Meyers et al., 1998). By flanking this resistance marker in the conditional construct with *FRT* sites, in addition to the *loxP* sites at either end of the gene essential sequence, it is possible to remove neomycin in ES cells by Flpe mediated recombination.

1.5 RECOMBINEERING

Traditional vector construction approaches, especially for the generation of conditional gene targeting vectors, have a number of drawbacks. As these methods rely on restriction and ligation of DNA fragments, the generation of a vector construct is often limited by the availability of appropriate and unique restriction sites. Additionally, some manipulations require constructs which are too large to be efficiently cloned by standard techniques, or their size cannot be tolerated by conventional cloning vectors.

The development of a phage-based homologous recombination system in *E.coli* – called Recombineering – has helped to overcome these problems. This method significantly simplifies the generation of transgenic and knockout constructs, as well as permitting the manipulation of large genomic DNA fragments like those present on Bacterial Artificial Chromosomes (BAC). As this method relies on subcloning by homologous recombination, it negates the need for unique restriction sites (which may be more difficult to find as length of the construct, to be manipulated, increases). This independence from appropriate endonuclease cleavage sites allows modifications to be made in any sequence.

Previously, cloning by homologous recombination was only done in yeast (Baudin et al., 1993), and was not possible in *E.coli*. However, being able to clone by homologous recombination in *E.coli* cells would have certain advantages over using the approach in yeast. Manipulating recombinant DNA in yeast is difficult and time consuming, and it often involves the transfer of the construct to *E.coli* cells for further processing (Bhargava et al., 1999). In contrast, recombinant DNA generated in *E.coli* can be used directly without the need not be transferred.

1.5.1 Phage-encoded recombination systems

Linear double-stranded DNA is unstable in *E.coli* cells due to the activity of the endogenous RecBCD exonuclease, thus preventing homologous recombination in this system (Cosloy and Oishi, 1973; Wackernagel, 1973). By studying the effect of defined mutations in the *E.coli rec* genes (Cosloy and Oishi, 1973; Lloyd and Buckman, 1985; O'Connor et al., 1989; Yang et al., 1997), researchers managed to

develop a bacterial recombination system which allowed linear dsDNA sequences with 40-50bp flanking homology to be integrated into a target BAC or plasmid present in *E.coli* cells (Muyrers et al., 1999; Zhang et al., 1998).

1.5.2 Rac-encoded RecET system

Francis Stewart and his colleagues were the first to show efficient integration of a PCR fragment flanked by 42bp homologies into the intended target site in a plasmid, by electroporation of the fragment into *recBC sbcA E.coli* strains (Zhang et al., 1998). sbcA is a mutation which activates the expression of *recE* and *recT* genes encoded in the cryptic Rac prophage, present in *E.coli* K12 strains (Clark et al., 1984). The expression of both these *rec* genes facilitates the homologous recombination of linear dsDNA sequences with only short homology arms (Zhang et al., 1998).

This discovery led to the generation of a pBAD-ETγ plasmid (Zhang et al., 1998); which, if transfected into *E.coli* cells, permitted these cells to perform homologous recombination.

The pBAD-ETγ plasmid contains the *recE* gene under the control of an arabinose-inducible promoter; as well as the *recT* and bacteriophage λ-*gam* gene, each expressed from a constitutive EM7 and Tn5 promoter, respectively. Upon induction with arabinose, recE is expressed. This protein, in turn, acts together with recT to allow homologous recombination in the *E. coli* cell, whilst expression of the gam protein gene inhibits the RecBCD-dependent degradation of the linear targeting cassette.

1.5.3 λ- encoded red system

In parallel to the RecET system, another set of recombination genes - encoded by bacteriophage λ - was utilised for recombineering in *E.coli* (Datsenko and Wanner, 2000; Yu et al., 2000). In the λ-phage, the recombination genes are called *redα* (or *exo*) and *redβ* (or *bet*). Both, *exo* and *bet* are analogous in their function to the *recE* and *recT* genes, respectively.

Exo is an exonuclease which progressively digests the 5'-end strand of dsDNA, leaving 3'-overhangs (Carter and Radding, 1971); whilst the beta protein binds to the single-stranded DNA ends and promotes annealing of the complementary sequences (Takahashi and Kobayashi, 1990). As in the RecET system, the gam protein, also present in the phage, inactivates the exonuclease function of RecBCD (Poteete, 2001; Stahl, 1998).

Similar to the ET cloning system, homologous recombination can be induced in *E.coli* cells carrying BACs or plasmids for modification, if these cells are transfected with a pBAD-αβγ vector (Datsenko and Wanner, 2000). In the pBAD-αβγ vector, *exo* is under the control of the arabinose inducible promoter, whilst *bet* is expressed from the constitutive promoter EM7, and *gam* from the Tn5 promoter.

E.coli strains expressing the red recombination system show very high levels of recombination activity, and BAC recombination efficiency in these cells was one- to three-fold better compared with *E.coli* cells containing the pBAD-ETγ plasmid.

Contrary to recombination in yeast cells, the presence of the arabinose-inducible promoter in both these systems controls the amount of time during which the recombination function is active in *E.coli* cells; thereby, reducing the chance of any unwanted recombination events occurring.

Even more control over the expression of the Red recombination genes was made possible with the development of the defective λ-prophage (Yu et al., 2000). In this prophage based system; *exo*, *bet* and *gam* are expressed from a defective prophage which is integrated into the *E.coli* chromosome. All three recombination genes in this system are expressed from a strong promoter (λP_L); which, in turn, has been placed under the control of a temperature sensitive λ-cI857 repressor. At 32ºC the repressor is active, and none of the red genes are expressed; however, a temperature shift to 42 ºC for 15 minutes leads to the inactivation of the repressor, resulting in very high levels of recombination activity.

The expression of the red recombination genes from the defective prophage, instead of the pBAD vector, has several advantages. For one, the presence of the repressor allows the expression of the red genes to be tightly controlled. This is particularly important as constitutive expression of *gam* can be toxic to *E.coli* cells, and cause plasmid instabilities (Sergueev et al., 2001). Furthermore, due to its

integration into the *E.coli* chromosome, the prophage is genetically stable in these cells and does not have to be maintained by drug selection.

The defective prophage has been integrated into DH10B *E.coli* cells, which are a commonly used host strain for BACs. The presence of the Red recombination system in the modified DH10B strain, called DY380, facilitates modification of BAC vectors, which can be transfected into these cells with an efficiency of 10^{-6} to 10^{-4} per electroporated cell (Lee et al., 2001).

The presence of the BAC, in such a recombinogenic cell line, facilitates a multitude of modification to be undertaken with relative ease. BAC modifications, possible in this system, range from precise base pair changes, to deletions and insertions of large stretches of DNA, up to 70kb or more (Copeland et al., 2001).

1.5.4 Generation of conditional targeting vectors

BAC recombineering has also been used for the construction of conditional targeting vectors (Liu et al., 2003). As recombineering allows precise restriction enzyme-independent DNA manipulation, *loxP* or *FRT* sites can be efficiently placed in any location within the targeting construct.

In this vector construction approach, *loxP* sites are integrated into the conditional construct in the form of a neomycin resistance cassette flanked by *loxP* and *FRT* sites. Efficient targeting of these constructs is facilitated by flanking the *FRT/loxP*-neo cassette with sequences homologous to the DNA sequences to be targeted. The neomycin resistance marker present in the cassette is under the control of both; a PGK promoter, and an EM7 promoter; allowing the gene to be expressed in *E.coli* and ES cells, respectively. Consequently, *E.coli* and ES cells can be selected for the presence of the cassette by either kanamycin (*E.coli*), or G418 (ES cells).

The generation of conditional targeting vectors by recombineering has been assisted further by the generation of DY380 *E.coli* cells carrying, either a *cre* recombinase gene in EL350 cells, or a *flp* recombinase gene in EL250 cells (Lee et al., 2001; Liu et al., 2003). Both strains are able to mediate homologous recombination and site-specific recombination. The former of these functions is controlled by temperature, whilst the latter can be induced by arabinose. This dual

regulation facilitates both; the integration of the *loxP/FRT* flanked targeting cassette, and the subsequent removal of the neomycin resistance marker from the targeted locus.

The use of BACs for the creation of conditional targeting vectors has a number of advantages. Due to the size of the average BAC (~200kb), any modified sequence in this construct will be flanked by very long regions of homology to the target locus in the mouse genomic DNA. This means that homologous recombination in ES cells occurs at relatively high frequencies (on average 3.8% in one study (Valenzuela et al., 2003)) without the need for positive or negative selection. The same study also showed that, due to these long stretches of homologous sequence; BAC vectors did not require isogenicity to the mouse ES cell line they are integrated into.

1.5.5 Cloning DNA by Gap repair

However, using BAC vectors to generate mutant conditional alleles in ES cells complicates the screening process for correctly targeted clones; as such long homology arms make Southern blot analysis, to determine the integration site, impossible.

Instead, homologous recombination in *E.coli* cells can be used to subclone a smaller genomic region of the target gene into a plasmid backbone. This subcloning approach, called Gap repair (or Gap retrieval), involves the homologous recombination of the free ends of a linear plasmid with the homologous sequences carried on the BAC (Lee et al., 2001; Liu et al., 2003; Zhang et al., 2000).

1.5.6 Generation and use of targeting vectors without a selectable marker

With the mouse and human genome fully sequenced, the focus for researchers now lies in annotating a function to all of the approximately 28,000 identified genes (Auwerx et al., 2004). Of particular interest is the replication of human diseases in the mouse model organism. As most human genetic defects

involve point mutations or deletions in a gene, efficient recreation of these mutations in the mouse gene orthologue is crucial.

BAC targeting vectors - containing point mutations or deletions in a gene of interest - can be rapidly generated using either a two-step SacB selection and counter-selection approach, or a single-step recombination of short double-stranded (dsDNA) or single–stranded DNA (ssDNA), into the sequence to be mutated (Muyrers et al., 2000). This two-step method relies on the integration of a selectable SacB-neo cassette, flanked by target specific homology arms, into the BAC vector. A subsequent targeted integration - with the mutated gene sequence of interest - leads to a replacement of the SacB cassette. The recombinant clones, carrying the designed gene deletion or point mutation, can easily be identified by counter-selection against the *SacB* gene from *Bacillus subtilis*, which renders *E.coli* cells sensitive to sucrose (Bramucci and Nagarajan, 1996).

The single-step method for introduction of sequence modifications into a gene of interest relies, predominantly, on the high level of homologous recombination in DY380 cells. Short dsDNA sequences, flanked by only 24bp of homology arms, have been shown to integrate into a target locus with an efficiency of 0.1 - 0.2% (Lee et al., 2001; Yu et al., 2000). This targeting efficiency can be improved by using ssDNA to generate the mutant vector construct, as ssDNA has been shown to integrate into the target sequence with an efficiency ranging from 1% to 6%, depending on the length of the flanking homology arms (Ellis et al., 2001; Swaminathan et al., 2001). As the final modified vector construct does not contain a selectable marker, and may only differ from the wild type target gene sequence by 1 base pair, large numbers of recombinant clones need to be screened by PCR. Due to the high number of recombinants generated per electroporated cell, when using the red recombination system, PCR based screening is a feasible method.

The relatively high targeting frequency of BAC vectors (Valenzuela et al., 2003) means that sequences containing only subtle mutations, without selectable markers, can be efficiently integrated into the ES cell genome. If conventional sized targeting vectors are used to integrate a sequence containing a subtle mutation into ES cells, a “tag and exchange” or “plug and socket” approach (see section 1.4.6.2) may be necessary.

1.5.7 High-throughput recombineering

The development of the recombineering techniques in *E.coli* has greatly improved the scope of possible DNA sequence modifications, without some of the limitations posed by previous approaches for targeting vector construction. Recombineering also presents a less complicated and faster approach.

Recent improvements of this method have mainly focused on adapting it for high-throughput construction of targeting vectors. Two research groups demonstrated the successful generation of more than 100 targeting vectors in a fast, efficient and almost fully automated 96-well format (Chan et al., 2007; Valenzuela et al., 2003). In both approaches, 96 different genes were chosen for simultaneous modification. For this purpose, DH10B cells harbouring BACs - which contained the genes of interest - were seeded into a 96-well plate. These cells were then made recombineering-competent by introduction of the *red* recombination genes, present in either a plasmid construct or the defective λ prophage (Chan et al., 2007; Valenzuela et al., 2003).

The BAC construct, in these recombineering competent cells, can then be efficiently modified by homologous recombination of PCR-amplified targeting cassettes, in a similar manner to that previously described for the generation of a conditional targeting construct.

Another bottleneck for the use of gene targeting as a high-throughput approach for mouse mutagenesis lies in the generation and identification of correctly targeted ES cell clones. The homologous recombination of traditional gene targeting vectors, harbouring the desired gene modification (such as deletions, f*lox*ed exons or reporter construct integrations) flanked by 2-6kb homology arms, into the ES cell genome has been shown to occur at a frequency of only 0.1 to 1%. This low level of homologous recombination coupled with the use of Southern blot screening methods to identify the correctly targeted ES cell clones make this stage of the gene targeting protocol very time consuming.

In order to speed up this process, Valenzuela and his colleagues opted to use BAC targeting vector constructs, which in their hands can integrate into the mouse genome with an average targeting frequency of 3.8% (Valenzuela et al., 2003). Furthermore, these researchers have developed a qPCR-based ES cell screening

approach which can greatly minimise the time taken to identify correctly targeted ES clones. This approach can be used to screen for gene targeting events designed to delete part of the wild type allele sequence. The qPCR primers used in this screening method are designed to the wild-type sequences to be deleted by the BAC targeting vector integration. This screen is based on measuring the copy number of the sequence which would be deleted upon vector integration, and comparing it to a control autosomal gene sequence. Therefore, the amount of PCR product produced from the sequence of interest (if it is present in two copies in the genome) should in wild-type cells be comparable to the amount of PCR product from an autosomal control gene after the same number of amplification cycles. Deletion of the target sequence in one allele would, therefore, result in a reduction of the PCR product, which correlates with only half the PCR template being amplified. Loss-of-native allele screening represents a suitable approach for the isolation of BAC-based gene targeting events as these vectors rarely get incorporated into the mouse genome by non-homologous, random integration. A further advantage of this method is that it can also be used for subsequent genotyping of mice carrying the BAC-targeted allele.

1.6 NEW TOOLS FOR LARGE SCALE MOUSE MUTAGENESIS

With the genomes of several animal models now fully sequenced, the new focus for researchers world-wide now lies in ascribing a function to all the identified protein-coding genes. Particularly in the mouse model, the most informative tool for the functional gene analysis is to deliberately engineer gene mutations. With a multitude of genes still needing to be analysed; new mouse mutagenesis methods are required, and existing ones need to be improved, to allow detailed gene functional analysis on a large scale. In recent years, advances have involved adding some of the advantages of other mutagenesis approaches to improve an existing technique.

As previously explained, the advantage of gene trapping has been its amenability to large scale mutagenesis in mice, allowing large numbers of mutant animals to be generated using just one vector construct. However, one drawback of this technique is that, unlike in conditional gene targeting approaches, timing and place of expression of the trapped gene cannot be controlled. Gene targeting, on the other hand, has the advantage of generating a defined genetic modification in a particular gene; as well as allowing control over the activation or silencing of a targeted gene, with the use of a conditional construct. But, as every targeting vector has to be generated specifically to a gene of interest, this approach lacks the possibility of high-throughput mutagenesis. According to current estimates and statistics produced by the International Mouse Knockout Consortium, only 4,000 targeted knockout alleles have been produced, to date.

The development of the recombineering approach, coupled with the 96-well method for targeting vector construction and ES cell screening (Chan et al., 2007; Valenzuela et al., 2003), has drastically improved the use of gene targeting on a large scale. In comparison, improvements to the gene trapping approach have involved the generation of a multi-purpose gene trapping allele (Schnutgen et al., 2005). Apart from the traditional features of a gene trapping vector, this multi-purpose construct also contains strategically placed site-specific recognition sites; such as *loxP*, *FRT* and their mutant derivatives. This system of directional site-specific recombination, called FlEX, facilitates the successive inversion of the trapping cassette - from the sense to the anti-sense orientation on the DNA strand (Figure 1.10). In the sense orientation, the cassette leads to silencing of the trapped gene; whilst inversion of the

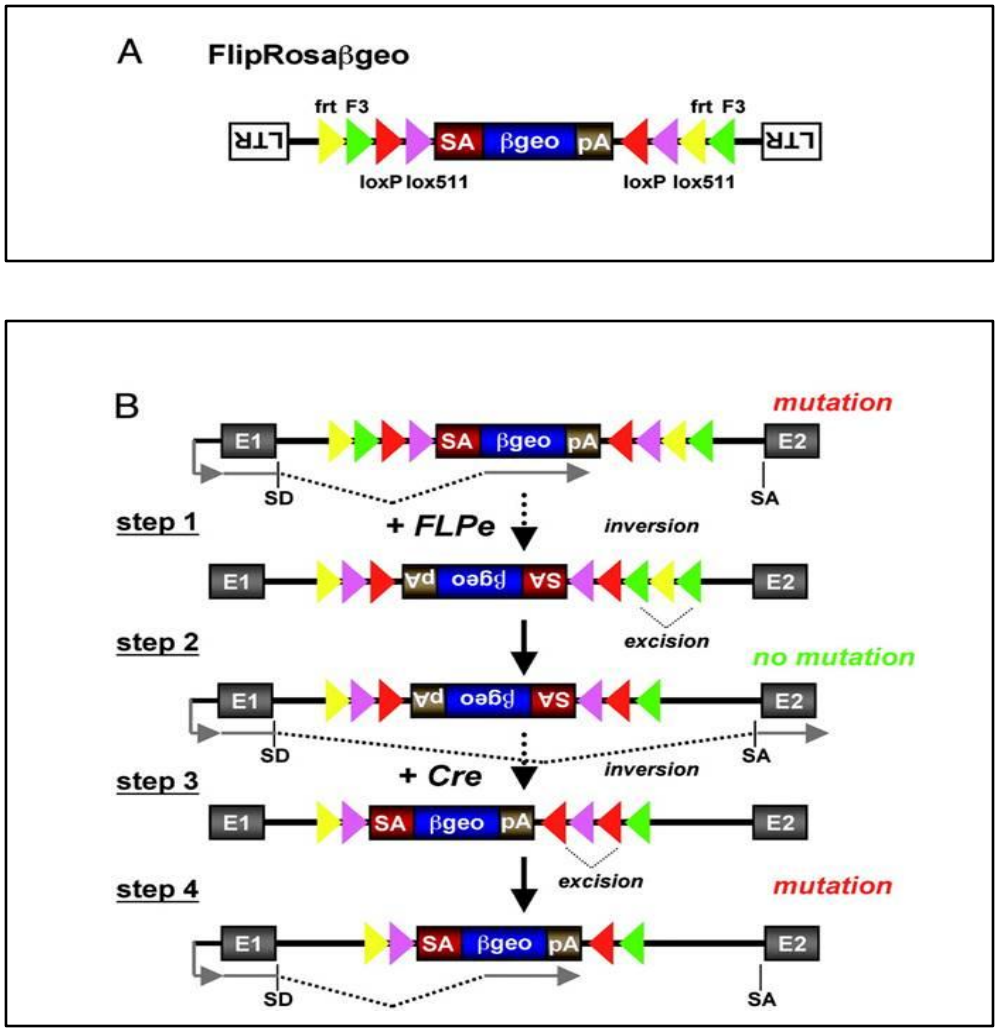


Figure 1.10: Conditional gene trap vector and mechanism of gene inactivation. (A) Schematic representation of the retroviral gene trap vector. This vector consists of two flanking long terminal repeats (LTR); two sets of heterotypic target sequences for the FLPe recombinase: *frt* (yellow triangles) and *F3* (green triangles) as well as heterotypic target sequences for the Cre-recombinase: *loxP* (red triangles) and *lox511* (purple triangles). The vector sequences at the core of this construct contain a splice acceptor sequence (SA); a *β-galactosidase-neomycin* (*βgeo*) fusion gene and a polyadenylation sequence (pA). (B) Conditional gene inactivation by the gene trap vector. The SA*βgeo*pA cassette flanked by recombinase target sites (RTs) in a FlEx configuration is illustrated after integration into an intron of an expressed gene. Transcripts (shown as gray arrows) initiated at the endogenous promoter are spliced from the splice donor (SD) of an endogenous exon (here, exon1) to the SA of the SA*βgeo*pA cassette. This leads to the *βgeo* reporter gene being expressed. The endogenous transcript, in this configuration, is captured and prematurely terminated at the cassette's pA, causing a mutation. In step 1, FLPe inverts the SA*βgeo*pA cassette onto the antisense, non-coding strand at either *frt* (shown) or *F3* (not shown) RTs and positions *frt* and *F3* sites between direct repeats of *F3* and *frt* RTs, respectively. By simultaneously excising the heterotypic RTs (step 2), the cassette is locked against reinversion because the remaining *frt* and *F3* RTs cannot recombine. This inversion reactivates normal splicing between the endogenous splice sites, thereby repairing the mutation. Cre-mediated inversion in steps 3 and 4 repositions the SA*βgeo*pA cassette back onto the sense, coding strand and reinduces the mutation. Abbreviations: *βgeo, β-galactosidase-neomycin* fusion gene; pA, polyadenylation sites; RT, recombinase target site; SA, splice acceptor; SD, splice donor; Figure adapted from Schnutgen *et al*. (2005).

cassette - to the anti-sense strand - allows reactivation of the trapped gene. These inversion events, controlling the expression of the gene, are driven by Cre and Flp recombinase.

The repertoire of current mouse mutagenesis approaches, for the large scale functional analysis of genes, is completed by: *Cre/loxP*-mediated conditional RNA interference (Ventura et al., 2004), and transposon-mediated saturation mutagenesis in the mouse germ line (Izsvak and Ivics, 2005; Takeda et al., 2007).

1.7 AIMS

The aim of this PhD project is to analyse the function of *Mospd1* and *Mospd3* in order to gain insight into the function of the *Mospd* family of genes. From the gene trap model of Mospd3, we know that loss of gene expression in *Mospd3*$^{Gt1lmf/Gt1lmf}$ mice causes neonatal death, and cardiac defects (specifically in the right ventricle). It was noted that, this phenotype closely resembled the defect associated with loss of desmosomal and adherens junction proteins, which provide structural integrity and contractile force transmission in the heart. In addition, the severity of the *Mospd3*$^{Gt1lmf/Gt1lmf}$ phenotype appeared to be highly dependent on genetic background, which indicated a possible genetic redundancy, maybe, involving another member of the Mospd family, such as Mospd1.

Although the gene trap model has helped to shed some light on the role of *Mospd3* in mammalian development; for a number of reasons, this mouse model has become unsuitable for further studies to elucidate the role of *Mospd* genes in heart development and function. These reasons include: the complexity of the gene trap integration; the incomplete correlation between expression of endogenous *Mospd3* and the *β-gal* marker gene; and, most importantly, the loss of severity of the neonatal death and cardiac malformation phenotype in *Mospd3*$^{Gt1lmf/Gt1lmf}$animals.

In order to gain further insight into the role of *Mospd* genes, particularly *Mospd3* and its closest family member, the aim of this PhD project was to generate new tools for these functional studies. These molecular tools include:

- **Conditional knockout alleles of *Mospd3* and *Mospd1***

With the aim of studying the effect of loss of function of *Mospd3* and *Mospd1* in a particular tissue, such as the heart, or at a specific time point of development; we decided to generate conditional knockout alleles of both genes. In addition to functional studies, the mice carrying the knockout alleles can also be used to generate double-knockout animals, in order to determine whether genetic redundancy exists between *Mospd3* and *Mospd1*.

- ***Mospd1* null ES cells**

In order to obtain evidence that *Mospd1* may also have a role in cardiac development, underpinning our hypothesis that it might be the genetic modifier affecting the *Mospd3*$^{Gt1lmf/Gt1lmf}$ phenotype, *Mospd1*-null ES cells were studied for their potential to differentiate into beating cardiac cells *in vitro.*

- **Custom antibodies specific to Mospd3 and Mospd1**

To study the hypothesis that Mospd proteins, in particular Mospd3, may be involved in the desmosome or adherens junction based structural integrity and force transmission in the heart; it is important to study the proteins localisation in the whole mouse model, as well as on a subcellular level. Antibodies are important analytical tools for such studies. As there were no commercial antibodies available to study the localisation and protein interaction of Mospd3 or Mospd1, we have generated custom antibodies, specific to either Mospd protein.

CHAPTER 2: MATERIALS AND METHODS

2.1 MOLECULAR BIOLOGY TECHNIQUES

All standard molecular techniques described in this chapter were carried out in accordance to the Molecular Cloning Laboratory Manual (Sambrook et al., 1989), unless otherwise stated. The same applies to the preparation of commonly used solutions for molecular techniques.

2.1.1 Transformation of Bacterial Cells

Frozen aliquots of 50µl chemically-competent DH5α (TOP10, Invitrogen) cells were thawed on ice. Once thawed, 1µl of plasmid DNA, or 5-20µl of ligation product, was mixed with the cells and incubated on ice for a further 20 minutes. The cells were then heat -shocked at 42ºC for 45 seconds and then left to recover on ice for 2 minutes. The transformation reaction was added to a 15ml Falcon containing 500µl SOC (Superoptimal culture broth) medium and incubated at 37ºC for 1 hour in an orbital shaker running at 200rpm. Finally, 50µl and 450µl, respectively, of this growth culture were spread on LB plates with the appropriate antibiotic and left to grow at 37ºC overnight.

2.1.2 Plasmid preparation (Miniprep and Maxiprep)

A single colony, picked from a LB or SOC agar plate or a 5ml starter culture, was grown in 5ml (Miniprep), or 200 to 400ml (Maxiprep), in LB or SOC medium and antibiotic overnight at 200rpm in an orbital shaker. The following day, the cells in culture were harvested by centrifugation at 6.000g for 5 minutes. The cells were then further processed to isolate plasmid DNA in accordance with the QIAgen kit and protocol for either Mini or Maxipreps.

2.1.3 Polymerase chain reaction (PCR)

This method was used for various purposes, such as: generating DNA fragments for subcloning, making Southern blot probes, or genotyping animals and cells. A representative PCR generally contained 10 to 100ng template DNA along with 10µM of 2 primers each in 1xPCR buffer supplemented with $MgCl_2$ (1.5mM),

dNTPs (10µM) and 1unit of Taq DNA polymerase (Invitrogen) in a final volume of 50µl. The template DNA was amplified in a HYBAID PCR-Express machine. A typical PCR proceeded as follows: 2 minutes denaturing at 94ºC, followed by 30-32 cycles of 94ºC for 15 seconds, 55ºC for 30 seconds, 72ºC for 1 minute, and a final extension cycle of 72 ºC for 10 minutes.

2.1.4 Restriction digest

The DNA was cut using the appropriate restriction enzyme and buffer. The restriction reaction was incubated for 2 hours to overnight at the temperature required for the enzymes activity (usually 37ºC). If restriction with two different enzymes was performed, the restriction buffer allowing optimal activity of both enzymes was used. The restriction fragments were then visualised and separated by gel electrophoresis.

2.1.5 Sequencing

Plasmid DNA samples were submitted for automated sequence analysis by the University of Edinburgh's School of Biological Sciences Sequencing Service. For this, 5µl of DNA obtained from Miniprep samples were combined with 1µl of 10mM uni-direction primer. Sequencing reactions were carried out according to the manufacturer's specifications, using the ABI PRISM Big-dye terminator premix sequencing kit (Perkin-Elmer). Samples, subjected to thermocycle-sequencing, were run on an ABI sequencer (Applied Biosystems).

2.1.6 DNA extraction from animal tissue and ES cells

Genomic DNA was extracted from animal tissues by boiling a small amount of the tissue for 20-30 minutes in 200µl of 0.05M NaOH solution. The lysis reaction is then neutralised by 25µl of 1M Tris (pH7.7), vortexed, and the cell debris pelleted at the bottom of an 1.5ml Eppendorf. The lysis supernatant contained genomic tissue DNA of good enough quality to be used in PCRs. For higher quality genomic DNA, tissues were lysed and DNA extracted using a DNeasy® Tissue kit (Qiagen).

For genomic DNA extraction from ES cells grown in 6-well plates or 25cm^2 flasks, cells were first harvested in a 15ml Falcon tube and then incubated overnight at 55ºC in 2 ml of lysis buffer (4M Urea, 10mM CDTA, 0.5% sarcosyl, 0.1M Tris pH8 and 0.2M NaCl) with 1mg/ml Proteinase K. The following day, 2ml of isopropanol was added to the lysis reaction and the tube was gently vortexed until a precipitate of genomic DNA was visible. The DNA was then spooled onto a sterile glass rod and resuspended in 500µl TE.

To extract genomic DNA from ES cells grown on a 96-well plate, the cells were first washed with 100µl of PBS, and then 50µl of ES cell lysis buffer (10mM Tris pH7.5, 10mM EDTA, 10mM NaOH, 0.5% sarcosyl) and 1mg/ml Proteinase K was added to each well. The plate was incubated in a moist chamber at 55ºC overnight. The following day, the DNA in the wells was precipitated by 200µl of 7.5mM NaOH in 95% ethanol. After 4 to 12 hours incubation at room temperature, the supernatant from the precipitation was tipped out of the plate, and the wells were washed 3 times with 100µl of 70% ethanol and then air-dried for 30 minutes.

2.1.7 DNA preparation for electroporation

20µg of targeting vector used for electroporation into ES cells were linearized with the appropriate enzyme, overnight. The restriction reaction was subsequently purified using an equal volume of 50:50 phenol: chloroform solution. The mixture was vortexed, spun at full speed in an Eppendorf Centrifuge 5415D microfuge, and the aqueous phase removed. A further purification using an equal volume of chloroform, vortexing and spinning followed. The purified vector was then precipitated with ammonium acetate and isopropanol, washed with 70% ethanol, and resuspended in 20µl tissue grade PBS, ready for electroporation.

2.1.8 Southern blot

Genomic DNA (10µg) was digested with the appropriate enzyme overnight at a temperature specific to the enzyme. Afterward, the digested DNA was precipitated with ammonium acetate and isopropanol, and resuspended in 30µl of TE.

When using genomic DNA isolated in a 96-well format, a restriction cocktail containing 1x buffer, 1mM spermidine, 100µg/ml RNAse and 10units of restriction enzyme per sample was used to digest the DNA. Of this restriction cocktail 30µl were added to each well with a multi-channel pipette, and the 96-well plate was incubated in a moist chamber overnight at the appropriate temperature.

The restriction fragments were resolved on a 0.7% agarose gel overnight. A picture of the gel was captured using a UVdoc system (Uvitec) with a ruler next to the DNA ladder, to allow future sizing of the DNA fragments. The gel was incubated for 5 minutes in depurination solution (10.25ml HCl in 500ml H_2O), and denatured for 45 minutes in 0.5M NaOH and 1.5M NaCl. After a further 15 minutes incubation in 10xSSC neutralisation solution, the DNA from the gel was transferred to an Hybond N+ membrane (Amersham) by wet Southern blotting capillary transfer, as described in (Sambrook et al., 1989). Following overnight blotting, the location of the wells was marked on the membrane. The membrane was rinsed in 2xSSC and air-dried. Using a UVdoc System (Uvitec), the DNA was cross-linked to the membrane.

2.1.8.1 Radioactive labelling and purification of probes

All radioactive work was performed under strict adherence to *Working with Radioactivity* guidelines for the university. Procedures involving radioactivity were performed in a designated room, and great care was taken to avoid radioactive spillage and contamination of work surfaces and materials.

The probe DNA was generated by high-fidelity PCR of the DNA from the appropriate BAC. Sequences of primers used for Southern blot probe generation can be found in Appendix table 2. The neomycin probe was generated by *NcoI* digestion of plasmid PL451 (Appendix Figure 3). The resulting PCR or restriction fragments were gel purified, and 100ng of purified DNA was added to nuclease-free water to a final volume of 11µl, and incubated at 100ºC for 5 minutes. This denatured DNA solution was snap-cooled on ice; and 4µl of HighPrime mix (Roche) containing 1 unit/µl Klenow polymerase, as well as 125µM dATP, dGTP and dTTP (each) in 50% glycerol was added. Also on ice, 5µl of radioactively labelled [α-^{32}P] dCTP (10mCi/ml) was added to the labelling mix, and the reaction was incubated at 37ºC for 1 hour. After this time the reaction mix was applied to a 1.5ml Sephadex G50 (Sigma) column. The column was washed with 1ml dH_2O and the labelled probe,

eluted in the 800µl of dH_2O, subsequently added. Finally, the probe DNA was denature for 5 minutes at 100ºC and hybridised to the Southern blot membrane.

2.1.8.2 Prehybidisation and hybridisation

ExpressHyb solution (Clontech) was pre-warmed to 65ºC in a water bath. The membrane was pre-hybridised for 30 minutes at 65ºC, in an evenly rolling Techne hybridisation bottle, in a Techne Hybridiser HB-1D oven. The denatured radioactively labelled probe was added to the 10ml pre-hybridisation solution, and the blot was further incubated at 65ºC overnight.

2.1.8.3 Washing

The following day, the hybridisation solution was discarded, in accordance with radioactivity disposal guidelines, and the blot was washed 3 times with 100ml of pre-warmed 2xSSC/ 1%SDS for 15 minutes at 65ºC. After a further 3x15 minute washes with 0.2xSSC/ 1%SDS at 65ºC, the blot was mounted in Saran-wrap in an X-ray cassette and exposed to X-ray film (Kodak) at -80ºC for 2 to 10 days. During the washing procedure, the level of background radioactivity was monitored with a Geiger counter; and when only a low localised radioactive signal was detected, the final washes were stopped.

2.2 TARGETING VECTOR GENERATION AND RECOMBINEERING IN *E.COLI*

2.2.1 Bacterial Strains and Plasmids

The *E.coli* strains, used in this study, were derived by transferring a defective lambda prophage into DH10B cells, to create DY380 cells (see Table 2.1 adapted from (Liu et al., 2003)). Addition of an arabinose-inducible cre gene (pBAD–cre) into the lambda prophage in DY380, led to the creation of EL350 cells (Lee et al., 2001). Chemically competent DH5α cells (Invitrogen) were used for the initial construction steps of the targeting vector.

The plasmids used in this study include: pNEB-DT, which carried a diphtheria toxin cassette for negative selection of targeting vector integrations into ES cells; as well as PL452 and PL451, both carrying a neomycin resistance gene flanked by either two *loxP* sites (PL452) or two *FRT* sites followed by a downstream *loxP* site (PL451). A list of plasmid features can be found in Table 2.1, and plasmid maps can be found in the appendix (Appendix figures 1, 2 and 3).

2.2.2 Subcloning of DNA targeting fragment from BAC into pNEB-DT vector

2.2.2.1 Preparation of BAC DNA

The *E.coli* cells containing the bacterial artificial chromosomes (BAC), used for the recombineering strategy, were obtained from a C57BL/6-BAC library, held at the BACPAC Resources Facility at the Children's Hospital Oakland Research Institute (CHORI). The BACs, carrying the entire genomic sequences of either *Mospd1* or *Mospd3*, correspond to the following NCBI accession numbers and CHORI-BAC clone number: *Mospd1*-BAC (AC100311 and RP23-123K17), *Mospd3*-BAC (BH063983 and RP24-129E14).

E.coli cells containing BACs were grown overnight in 600ml LB broth supplemented with 20µg/ml chloramphenicol. The cells were collected, and DNA was extracted using a QIAGEN Maxiprep kit. In accordance with the instructions for Plasmid DNA purification from very low-copy plasmid/cosmid containing cells, 20mls of buffers P1, P2 and P3 were used. The lysate, containing the cell's protein

Table 2.1: Recombineering reagents.

Strains	Genotype
DY380	DH10B [λc1857 (cro-bioA < >tet]
EL350	DH10B [λc1857 (cro-bioA < >araC-P_{BAD}*cre*]
Selection cassettes	
PL451	*FRT-PGK-EM7-NeobpA-FRT-lox*
PL452	*loxP-PGK-EM7-NeobpA-loxP*
Other plasmids	
pNEB-DT	pNEB193-*DT*

precipitate, was spun twice at 17,900xg for 30 minutes. The clear, protein precipitate-free supernatant was applied to 2x500 Qiagen MaxiPrep column, and after washing with buffer QC, the DNA was eluted with 55ºC elution buffer QF. The DNA was then precipitated with isopropanol, and washed with 70% ethanol, and subsequently the two samples of the same clone were resuspended and pooled in 300µl of TE.

2.2.2.2 Subcloning of *Mospd1* by restriction-ligation methods

The *Mospd1*-BAC (AC100311) was digested with *SmaI* and *SacI* enzyme, and the resulting restriction fragments were separated by gel electrophoresis on a 0.7% gel. The bands, corresponding to a size of approximately 12kb, were cut from the gel and re-cast in a fresh 0.7% gel, to be further resolved at 20V, overnight. In parallel, the pNEB-DT vector (Appendix figure 1 and Table 2.1) was digested with *SmaI* and *SacI* restriction enzyme. The linearized pNEB-DT plasmid DNA was then precipitated with ammonium acetate, resuspended in 50µl 1xCIP (Calf Intestinal Alkaline Phosphatase) buffer, and heated to 65ºC for 5mins, and then placed on ice. After the addition of 1µl CIP (Calf Intestinal Alkaline Phosphatase, Roche), the reaction was incubated at 37ºC for 2hrs, and then stopped by heating to 65ºC. Like the *SmaI/SacI* digest of the BAC, the linearized pNEB-DT plasmid was purified twice on a 0.7% gel. The plasmid and the BAC fragments were purified using QIAexII kit (Qiagen). Finally, 1.5µg BAC fragment was mixed with 0.5µg of linearized pNEB-DT in 1x ligation buffer and 1µl T4 ligase (Invitrogen). The reaction was incubated in 37ºC for 2hrs, resulting in the directional ligation of the BAC (*SmaI/SacI*) fragment into the *SmaI/SacI* linearized pNEB-DT vector. This targeting vector was then transformed into DH5α cells (Section 2.1.1).

2.2.2.3 Cloning of homology regions for Mospd3 gap retrieval into pNEB-DT

Two genomic regions of the *Mospd3*-BAC, flanking the 13kb *Mospd3* DNA fragment, were PCR amplified using primer pairs: Mospd3A'-HindIII and Mospd3B'-BamHI, or Mospd3Y_3-BamHI and Mospd3Z_3AvaI. Primer sequences are listed in Appendix table 1. These primers amplify the following fragments: a 204bp (A,B) fragment with the primer pair one, and a 207bp (Y,Z) fragment with primer pair two.

PCR amplification (Invitrogen, Platinum Taq Polymerase High Fidelity kit) was performed with the following settings: 94°C for 5min, then 30 cycles of 94°C for 15sec, 55°C for 30sec, and 72°C for 1min, followed by 72°C for 10min. The PCR products were then digested with *HindIII* and *BamHI* for fragment A,B; and *BamHI* and *AvaI* for fragment Y,Z. Subsequently, 10μl of each fragment were simultaneously ligated with 1μl *AvaI/HindIII*-linearized pNEBDT vector. To facilitate gap retrieval of the 13kb *Mospd3* fragment for the targeting vector, the pNEB-DT vector, containing the *Mospd3* homology arms A,B and Y,Z; was electroporated into recombination efficient EL350 cells containing the *Mospd3* BAC (BH063983) as described in section 2.2.2.5.

2.2.2.4 Electroporation of BAC or Plasmid DNA into EL350 cells

EL350 cells were grown in 5ml LB/ 2% sucrose broth overnight at 32°C. The cells were collected the next day by centrifugation for 10 minutes at 3,000xg at 4°C. The cells pellets were resuspended in 900μl of ice-cold water, transferred to a 1.5ml Eppendorf tube (on ice), and centrifuged at 16,000xg using a bench-top Heraeus Sepatech Biofuge13 for 25sec at 4°C. The supernatant was removed and cells resuspended in 900μl fresh ice-cold water. The process was repeated 5 more times. Finally, the cell pellet was resuspended in 50μl ice-cold water and transferred into to pre-cooled electroporation cuvette (0.1cm gap). Next, 4μl *Mospd3*-BAC (300ng) or 2μl pNEB-DT-*Mospd1* plasmid DNA (300ng) was added and mixed. Electroporation was performed using a BIO-RAD electroporator under the following conditions: 1.75kV, 25μF with pulse controller set at 200Ω. After electroporation, 0.5ml SOC was added to each cuvette, and then transferred with the electroporated EL350 cells into a 1.5ml Eppendorf tube, and grown at 32°C for 1h. Cells were spread on plates with either chloramphenicol (*Mospd3*-BAC) or ampicillin (pNEB-DT-*Mospd1*).

2.2.2.5 Gap retrieval

EL350 cells containing *Mospd3*-BAC (BH063983) were inoculated into 5ml of LB broth with 20μg/ml chloramphenicol and grown overnight at 32°C with shaking. The next day, 1.0ml of overnight culture was transferred to 20ml of super optimal broth with glucose (SOC) and incubated for approximately 2.5h with

shaking, until the OD600 reached 0.7 to 0.8. Then 10ml of the cells were transferred to a new flask and shaken at 42°C for 15min, whilst the remaining 10ml of culture were put on ice to be used as un-induced control. After their 15min incubation, the induced cells were also put on wet ice and shaken to cool them rapidly. After 5min on ice, the induced and un-induced cells were transferred to chilled Falcon tubes and spun at 5,000rpm (4°C) for 5min. Cells were resuspended in 900µl ice-cold water, then transferred to a 1.5ml Eppendorf tube, and washed 5 times with ice-cold water, as described above. Finally, the cell pellet was resuspended in 50µl ice-cold water. This cell suspension was mixed with 2µl *BamHI* linearized pNEB-DT vector, containing the *Mospd3* homology regions AB and YZ, and the cells were electroporated as described above. The electroporated cells were plated on SOC agar containing ampicillin.

2.2.3 Introduction of the *neo* cassettes

The 2kb *loxP-neo-loxP* cassette was PCR amplified from the PL452 vector (Appendix Figure 2 and Table 2.1) using 80mer primers Mospd1-loxpneo-1_for and Mospd1-loxpneo-2_rev, along with the Invitrogen Platinum Taq Polymerase High Fidelity kit. The primers used in the case of *Mospd-3* were: Mospd3-loxpneo-1_for and Mospd3-loxpneo-2_rev. Expand High Fidelity polymerase enzyme (0.5µl, Roche), 1µl of each primer and 1µl template in 21.5µl water were mixed with 25µl Failsafe PCR mix J for the amplification. All sequences of primers used for recombineering are listed in Appendix table 1.

The approximately 2kb *FRT-neo-FRT-loxP* cassette was PCR amplified from PL451 vector (Appendix Figure 3 and Table 2.1) using 70mer primers Mospd1-FRTneo_p1 and Mospd1-FRTneo_p2 for *Mospd-1* targeting, and primers Mospd3-FRTneo_p1 and Mospd3-FRTneo_p2 for *Mospd-3*. Failsafe mix H and I were used for this PCR.

The PCR, in both cases, was performed using a two-stage program with the following settings: pre-warm at 94°C for 2min, after which the PCR was added to the machine; this was followed by 94°C for 2min; then 10 cycles of 94°C for 15sec, 45°C for 30sec, and 68°C for 4min; followed by 20 cycles of 94°C for 15sec, 60°C

for 30sec, and 68°C for 4min - with an additional 5sec extension time after each cycle. The reaction was concluded by 68°C for 7min.

The PCR product was re-precipitated and digested for 1 hour with *DpnI* to eliminate plasmid template contamination. The *Dpn1* restriction enzyme only acts on methylated DNA, like that of the *E.coli* plasmid; but does not cut the unmethylated DNA of the newly synthesised PCR product. The PCR fragment DNA was further purified using a QIAexII kit (Qiagen). Finally, 300ng of the purified PCR product was electroporated into EL350 cells containing targeting vectors as described in the previous section 2.2.2.5. The electroporated cells were plated out on SOC plates containing kanamycin.

2.2.4 Excision of the f*lox*ed *neo* cassette

EL350 cells were inoculated into 5ml of LB broth with 2% sucrose, and grown overnight at 32ºC. The next day, 10ml of fresh LB were inoculated with 1 ml of EL350 overnight culture and grown at 32ºC until the culture reached O.D.=0.5. 100µl of L(+)-arabinose (Sigma) was added to the culture, and *Cre* induction from arabinose promoter (pBAD-cre) was allowed to proceed for 1 hour at 32ºC. The cells were collected at 3,000xg, and washed in ice cold water as described previously. Arabinose-induced EL350 cells, and un-induced controls, were electroporated, as before, with 10ng of the f*lox*ed neo cassette containing the targeting vector. The cells were plated, both, on ampicillin and kanamycin SOC plates.

2.3 CELL CULTURE AND MANIPULATION

In the course of this project, four different cell lines have been used: two feeder-independent embryonic stem (ES) cell lines, CGR8 and E14; as well as immortalised cell lines COS7, and mouse C57BL/6 keratinocyte cells (Millipore/ Chemicon).

Both ES cell lines were maintained in ES cell culturing medium, consisting of 1X Glasgow Minimal Essential Medium (Gibco) supplemented with 20% fetal calf serum, 0.25% sodium bicarbonate (Gibco), 0.1% non-essential amino acids, 2mM L-glutamine (Gibco), 1mM sodium pyruvate (Gibco), 0.1mM β-mercaptoethanol (Sigma) and 100 U/ml Leukaemia Inhibitory Factor (LIF). The working concentration of LIF, obtained in culture supernatant from COS7 cells transfected with murine LIF expression plasmid pDR10, corresponds to 100x the minimum concentration required to keep CP1 ES cells (Bradley et al, 1984) undifferentiated. COS7 cells were cultured in ES cell medium without LIF. The mouse C57BL/6 keratinocyte cells were cultured in the culture medium provided, for this purpose, by Millipore.

To ensure adherence of cells in the tissue culture flasks, ES cells were plated on tissue culture plastic pre-coated with 0.1% gelatine. COS7 cells and keratinocytes will adhere straight to tissue culture plastic, without any other matrix.

All cell lines were grown at 5% CO_2 in a 37ºC humidified incubator (Galaxy S, Wolf Laboratories).

2.3.1 Thawing cells

Cells, stored at -140ºC, were thawed rapidly in 37ºC water bath and transferred into a Universal tube containing pre-warmed culture medium to a final volume of 10ml. The cells were centrifuged at 180xg for 3 minutes, the supernatant removed. Then, the cell pellet was carefully resuspended in fresh culture medium and transferred into a 25cm^2 culture flask with 10ml culture medium. The culture medium was replaced after 4 to 12 hours.

2.3.2 Passage and expansion of ES cells

Cells were generally passaged every 2 days, when they had reached around 70% to 80% confluency. The medium was aspirated from flask, and the cells were washed with PBS and then incubated for 5 minutes in enough trypsin solution (1% trypsin, 1% chick serum and EDTA in PBS) to cover the bottom of the culture flask or plate. Once the cells had lifted off the tissue culture plastic and been agitated into single cell suspension, the trypsin was neutralised with at least 5X volume of culture medium in a Falcon tube, and cells were spun at 180xg for 5 minutes. The cell pellet was resuspended in 10ml fresh medium and cell number was determined using a haemocytometer on a Leitz Labovert microscope. $1x10^6$ cells per ml (in 10ml) were transferred to a new $25cm^2$ flask, or $3x10^6$ cells per ml (in 30ml) to a $75cm^2$ flask.

After 2 days of growth, cells from a $75cm^2$ flask can be trypsinised and passaged at approximately $8x10^6$ cells in 80ml of culture medium to a $225cm^2$ flask.

2.3.3 G418 concentration kill curve

Wild type ES cells were plated at $4x10^3$ cells into each well of a 6-well plate, and left overnight to settle in 5ml of ES cell medium with LIF. On the following day, the medium was replaced with 5ml of fresh medium containing different concentrations of G418 (Geneticin, PAA). The concentrations used were 0, 100, 150, 200, 250, 300, 350 and 400µg/ml. The medium containing G418 was replaced on a daily basis, and cells monitored. The optimal G418 concentration was defined as the minimum concentration used to kill all wild type ES cells. This was found to range from 260-280µg/ml.

2.3.4 ES cell electroporation

Cells grown on a $225cm^2$ flask were trypsinised, pelleted and resuspended in 20ml culture medium. Cells were counted; then, $3x10^7$ cells were centrifuged again and resuspended in 0.6ml PBS. This cells suspension was mixed with 20µg of linearized targeting vector DNA and then transferred into an electroporation cuvette

(Invitrogen, 0.4cm gap). The cells were electroporated in a GenePulser™ (Biorad) with one pulse of 250V, set at a capacitance of 500μF.

After a recovery period of 5 minutes the cells were plated onto 11 pre-gelatinised 100mm tissue culture dishes, and cultured in 10ml ES cell culture medium per plate.

On the following day, the medium on 10 of the plates was replaced with ES cell medium containing 280 to 320μg/ml G418 (PAA). The antibiotic selection continued for 6-9 days with daily washing and media changes. After this period, individual resistant colonies were picked into 96-well plates. Once the cells in the majority of the wells reached confluency, the cells were trypsinised and triplicate plated. After a further 2 days of growth, one of the triplicate plates was frozen for further expansion of correctly targeted clones. The cells in each well were trypsinised with 50μl of trypsin, and then agitated into single cells suspension on addition of 50μl of 20% DMSO in fetal calf serum. The plate was sealed and stored at -80ºC. The two remaining plates were left for a further one to two days. DNA was extracted from these plates (see section - DNA extraction from cells).

2.3.5 Transient Cre expression in ES cells

ES cells containing a *loxP* flanked allele were harvested from a 75cm^2 tissue culture flask and counted. 5x10^6 cells are resuspended in 0.6ml of PBS, mixed with 25μg *Cre* expression plasmid (pCAGGS-Cre-IRESpuro (plasmid map in Appendix Figure 4), and electroporated. The electroporated cells were returned to a 75cm^2 flask to recover overnight. The following day, the cells were trypsinised again, counted, and replated at varying cell numbers of 500; 1,000 and 2,000 cells in 10mm plates with ES cell medium. After 7-9 days of culture, 20- 30 distinct colonies were picked into a 96-well plate. These clones were subsequently processed as in section 2.3.4.

2.3.6 Transfection of DNA into COS7 cells

COS7 cells were plated at 2x10^5 cells per well in a 6-well plate. After 2 days, 5μl of Lipofectamine 2000 (Invitrogen) and 250ng vector DNA was added to 50μl of serum-free OptiMEM (Gibco), and incubated for 20 minutes at room temperature.

This vector/lipofectamine mix was added to each well of the 6-well plate, in which the ES medium had been replaced by 1ml OptiMEM. After 4 hours of incubation with this transfection mix (at 37ºC in a humidified 5% CO_2 incubator), the OptiMEM on the cells is replaced with fresh ES medium again, and cells are grown for a further 2 days before protein was extracted.

2.3.7 ES cell self renewal assay

$1x10^3$ ES cells were plated in each well of a 6-well plate with ES cell medium containing LIF. The following day, the wells were washed with PBS and the medium replaced. In half of the wells, LIF was added to the medium; whilst the other half did not receive any LIF in the medium. After a further 4 days under these culture conditions, the ES cell colonies were stained for alkaline phosphatase activity.

The cells were washed with PBS and the staining procedure was performed using an Alkaline Phosphatase Kit (Sigma-Aldrich). In accordance with the protocol, cells were first fixed (67% acetone, 25% citrate solution, 8% formaldehyde) for 30 seconds, then rinsed in water and subsequently incubated with the colour substrate solution for 15 minutes in the dark. After a final rinse and drying of the plate, the alkaline phosphatase staining of individual colonies was analysed using an inverted Zeiss Axiovert25 microscope.

2.3.8 Preparation of embryoid bodies (EB) and cardiomyocyte assay

ES cells were trypsinised from a $25cm^2$ flask, resuspended and counted. $6x10^5$ cells were added to a Universal tube in a volume of 20ml ES cell medium with LIF. Using a multichannel pipette, the lid of a square Petri-dish (measuring $120mm^2$) was covered with 10µl drops of this ES cell suspension. The lid, with these hanging drops attached, was then placed on the base of the square dish, filled with 10ml of water to maintain a humid atmosphere. The hanging drops were incubated at 37ºC in a humidified incubator for 2 days and then harvested by tapping the dishes on the bottom of the hood. The collected, 2 day old, embryoid bodies in the medium were spun at 800 rpm for 3 minutes, and fresh medium (without LIF, but supplemented with Penicillin/Streptomycin solution (Sigma) at 1:100) was added. The embryoid

bodies were then cultured in suspension, in a bacteriological 100mm dish, for a further 5 days, with medium changes every two days.

After 5 days, the embryoid bodies are transferred (one EB per well) into a pre-gelatinised 24-well plate. The embryoid bodies were analysed every 2 days, and the number of EBs with beating foci was counted.

2.3.9 Karyotyping of ES cell clones

ES cells were passaged at $2x10^6$ cells into a $25cm^2$ flask. The next day, 100µl of KaryoMAX (Invitrogen) was added to the flask and cells were incubated at 37ºC in the incubator for 2 hours. After that time, the cells were trypsinised, pelleted in 15ml Falcon tubes, and the medium removed. Under gentle vortexing, 8ml of a 0.56% KCl solution were added drop-wise to resuspend the pellet, and this cell suspension was incubated in a 37ºC water bath for 12 minutes. Next, 2ml of freshly prepared fixative (methanol 3:1 acetic acid) was added to the tube and mixed gently. The cells were then spun at 1000 rpm for 5 minutes, the supernatant removed and the cells resuspended in 8ml of fixative. This spin and resuspension process was repeated 3 times, and after the final spin the cells were resuspended in about 1ml of fixative. One drop of this final suspension was then dropped onto a slide, which had first been pre-cooled at 4ºC in ethanol and then dried.

These karyotyping spreads were stained with a 1:2,500 dilution of DAPI (Sigma) and a cover slide mounted. Images of 20-30 karyotyping spreads, for each clone, were captured under fluorescent light on a Zeiss Axioskop2 microscope using a 63x oil emulsion objective. The camera used was a ProgRes C14 from Jenoptik, and images were modified and saved using Open*lab*™ software. The number of chromosomes in each of the spreads was counted manually, using Adobe Photoshop, by marking each counted chromosome on the image with a red dot.

2.4 ANTIBODY PURIFICATION AND TESTING

2.4.1 Synthetic peptide generation and antibody production

Synthetic peptides corresponding to the N-terminal 15 amino acids of Mospd1 and Mospd3 were generated by Thistle Research, and also by Yorkshire Biosciences. These peptides were conjugated to a immunogenic KLH carrier protein. The synthetic peptides - generated by Thistle Research - were supplied to PTU/BS (Roslin), who raised polyclonal antibodies in two rabbits for each peptide. The animals were bled prior to inoculation with the peptides (pre-immune) as well as 4, 8 and 12 weeks after subcutaneous injection of the peptide antigen.

The synthetic peptides - generated by Yorkshire Biosciences - were used to inoculate mice. Splenocytes and lymph node cells from mice showing the best response to the antigen, tested by ELISA screening, were then fused to histocompatible myeloma cells. Fusion hybridomas were selected using HAT medium, and subsequently screened (using ELISA) for the production of the desired monoclonal antibodies. Three culture supernatants containing monoclonal antibodies were supplied for further analysis.

2.4.2 ELISA analysis of rabbit antisera

Dynatech Immulon 96-well ELISA plates were coated overnight with 0.5µg of synthetic peptide per well, at room temperature, in sodium carbonate-bicarbonate buffer. The plates were blocked for 30 minutes, at room temperature (RT), with 250µl/well blocking buffer (PBS with 0.05% Tween plus 2% (wt/vol) BSA), before being rinsed several times with distilled water. A 50 µl/well volume of rabbit antisera (pre-immune and week 4, 8 and 12); diluted in 1/20,000 increments from 1/20,000 to 1/160,000 with ELISA blocking buffer; was added to 3 wells each of the plate and incubated for 2 hours at RT. The wells of the plate were then washed 3 times with wash buffer (PBS with 0.05% Tween 20 and 0.25% BSA) before being incubated with 50 µl/well goat anti-rabbit IgG horseradish peroxidase conjugate (diluted 1/1,000 in blocking buffer) for 2 hours at RT. After being washed 3 times with wash buffer, the plate was then incubated with 100 µl/well of the OPD substrate (Sigma

p9187), for 2 hours at RT. The reaction was left to proceed in the dark for 30 minutes before the absorbance was measured at 450nm on a microplate reader (Vmax kinetic).

2.4.3 Antibody purification

Specific antibodies were purified from 8-week rabbit antiserum using an S-PAC column, containing the bound synthetic peptide (Thistle Research, Glasgow, UK), used for rabbit immunisation. The purification was performed in accordance with the protocol supplied by the manufacturer. The S-PAC90 column was equilibrated with 20ml of 2M KSCN solution, followed by 50ml of PBS, before being loaded with 2ml of ultra-filtered (0.45µm filter) rabbit antiserum (diluted 1/10 in PBS). Subsequently, the column was washed with 50ml PBS before eluting the bound antibody with 5ml of 2M KSCN solution. The eluate was collected in 2 fractions of 2.5ml sample each, and each fraction was immediately desalted using a PD10 column (Amersham). The column was pre-equilibrated with 25ml PBS prior to addition of the 2.5ml antibody eluate sample. The desalted antibody was subsequently eluted from the PD10 column using 3.5ml PBS, and 0.5ml elution fractions were collected. The amount of antibody in each of these fractions was measured using a spectrophotometer at 280nm.

2.4.4 Protein extraction from cells and animal tissues

Animal tissues were dissected and immediately placed in ice cold PBS, containing a 2x stock solution of Complete Protease Inhibitor Cocktail (Roche) and 20µg/ml of pepstatin (Roche). The tissues were then homogenised on ice using a Polytron Homogeniser. Finally, an equal volume of 2x Laemmli sample buffer (62.5mM Tris-HCl pH6.8, 1% SDS, 0.1M glycerol, 0.25% β-mercaptoethanol; and traces of bromophenol blue) was added to the homogenate and boiled in a hot block for 10 minutes.

Protein samples were obtained from cells in monolayer culture by adding 500µl of cold PBS/ protease inhibitor cocktail into a 6-well plate containing 90% confluent cell culture washed with PBS. Upon the addition of 500µl 2x Laemmli

sample buffer, the lysed cell were lifted of the well using a cell scraper and transferred to a 1.5ml Eppendorf. The cell lysate was boiled for 10 minutes at 100ºC.

2.4.5 TritonX100 protein extraction from ES cell clones

ES cells from monolayer (25cm^2 flasks) were harvested and washed in PBS. Cells were resuspended in 500µl TritonX100 lysis buffer (cold PBS/ protease inhibitor cocktail with EDTA (Roche) and 0.5% TritonX100). The cells were lysed at 4ºC for 30 minutes (on rolling platform), and the lysate was spun at 1,000xg for 30 minutes. The obtained supernatant was used for antibody preclearing.

2.4.6 Gel electrophoresis and Western blotting procedure

The proteins, contained in the solubilised cells and tissues, were separated by SDS-PAGE gel electrophoresis. The samples were denatured for 5 minutes at 80ºC, loaded onto a precast 10% or 12% Tris-HCl gel (Biorad), and a current of 70V to 80V was applied for 2 to 3 hours. The gels were then soaked in transfer buffer (Tris 3.03g, glycine 14.4g and 20% methanol made up in 1 litre) along with Nitrocellulose membrane (Whatmann) and 3MM Whatmann paper. Subsequently, the gel and membrane were assembled in a commercial semi-dry blotting apparatus (Biorad), according to manufacturer's instructions, and the proteins transferred onto the membrane for 1 hour at 15V.

Blots were blocked for non-specific binding for 1 hour at RT, or at 4ºC overnight, in 5% milk in TBS with 0.05% Tween20. The appropriate dilution of primary antibody in 5% milk/ TBS with Tween20 was added to the membrane and incubated for 2 hours at RT, or overnight at 4ºC, depending on the binding efficiency of the antibody.

Following three 10-30 minute washes, the membrane was incubated for 1 hour in secondary HRP-conjugated antibody, diluted in 0.5% BSA in TBS with 0.05% Tween (according to manufacturer's instructions). The membrane was washed a further 3 times to remove any unbound antibody. All incubations and washes were performed on a shaking platform.

After the final washes, the membrane was exposed to ECL chemi-luminescent substrate (Amersham Pharmacia Biotech) for 1 minute and visualised on photographic film (Kodak).

2.4.7 Antibody preclearing and peptide competition

Purified polyclonal antibodies were mixed with the preclearing substance (mouse serum (Sigma) or TritonX100 lysates of ES cells) at a ratio of 1:1. This mix was incubated at 4ºC overnight, on a shaking platform, before being used for primary antibody incubation of a western blot, at an appropriate dilution in 5% milk in TBS with 0.05% Tween20.

For peptide antigen competition, purified antibodies were mixed at a ratio of 2:1 with a 2mg/ml solution of peptide antigen in PBS and incubated at 4ºC overnight (on a shaking platform), before being used for primary antibody incubation of a western blot at an appropriated dilution in 5% milk in TBS with 0.05% Tween20.

CHAPTER 3: GENERATION OF THE *MOSPD1* AND *MOSPD3* TARGETING VECTOR AND CONDITIONAL ALLELE IN ES CELLS

3.1 INTRODUCTION

Mospd3 is one of four genes, of an as yet largely undescribed gene family, in the mouse model. The gene was discovered through a gene trap experiment aimed at identifying novel developmental genes. Loss of Mospd3 in mice carrying two trapped alleles of the gene resulted in cardiac defects as well as neonatal death (Pall et al., 2004). This phenotype closely resembled other mouse model phenotypes caused by deletion of genes involved in intercalated disc attachments and contractile force transmission in the heart. This similarity indicates that Mospd3 may also function to provide structural integrity in the heart.

Unfortunately, for a number of reasons, the gene trap model of *Mospd3* was unsuitable for further analysis of the genes function. These reasons include: the genetic complexity of the gene trap vector integration; the finding that the expression of the vectors *β-gal* marker gene did not fully correspond to the endogenous expression pattern of *Mospd3* in the embryo; and, most importantly, the fact that some aspects of the gene trap phenotype have been lost (Pall et al., 2004).

Nevertheless, loss-of-function approaches in the mouse model represent a particularly powerful experimental tool for studying genes involved in development. We, therefore, chose to generate a conditional knockout of *Mospd3*. The plan for this conditional allele involved flanking important 5' exons of *Mospd3* with *loxP* sites in such a way that Cre-mediated deletion results in a complete loss of the encoded Mospd3 protein. The combination of this *loxP*-flanked construct with tissue-specific and inducible Cre expression systems would make it possible to determine if *Mospd3* function is particularly required in the heart tissue. This system should also provide an answer to whether Mospd3 is required either for embryonic heart development, or the maintenance of cardiac structural integrity in newborn and adult mice.

In addition to *Mospd3*, we also chose to generate a conditional knockout mouse model of *Mospd1* in order to determine whether this member of the *Mospd* gene family is functionally redundant to *Mospd3*.

This chapter describes the production of conditional *Mospd1* and *Mospd3* targeting vectors by recombineering, and their subsequent introduction into two different mouse ES cell lines to produce mutant alleles for functional analysis. We also describe the generation of ES cells carrying a hemizygous null allele of *Mospd1*.

3.2 RESULTS

3.2.1 Generation of *Mospd1* targeting construct and targeted allele

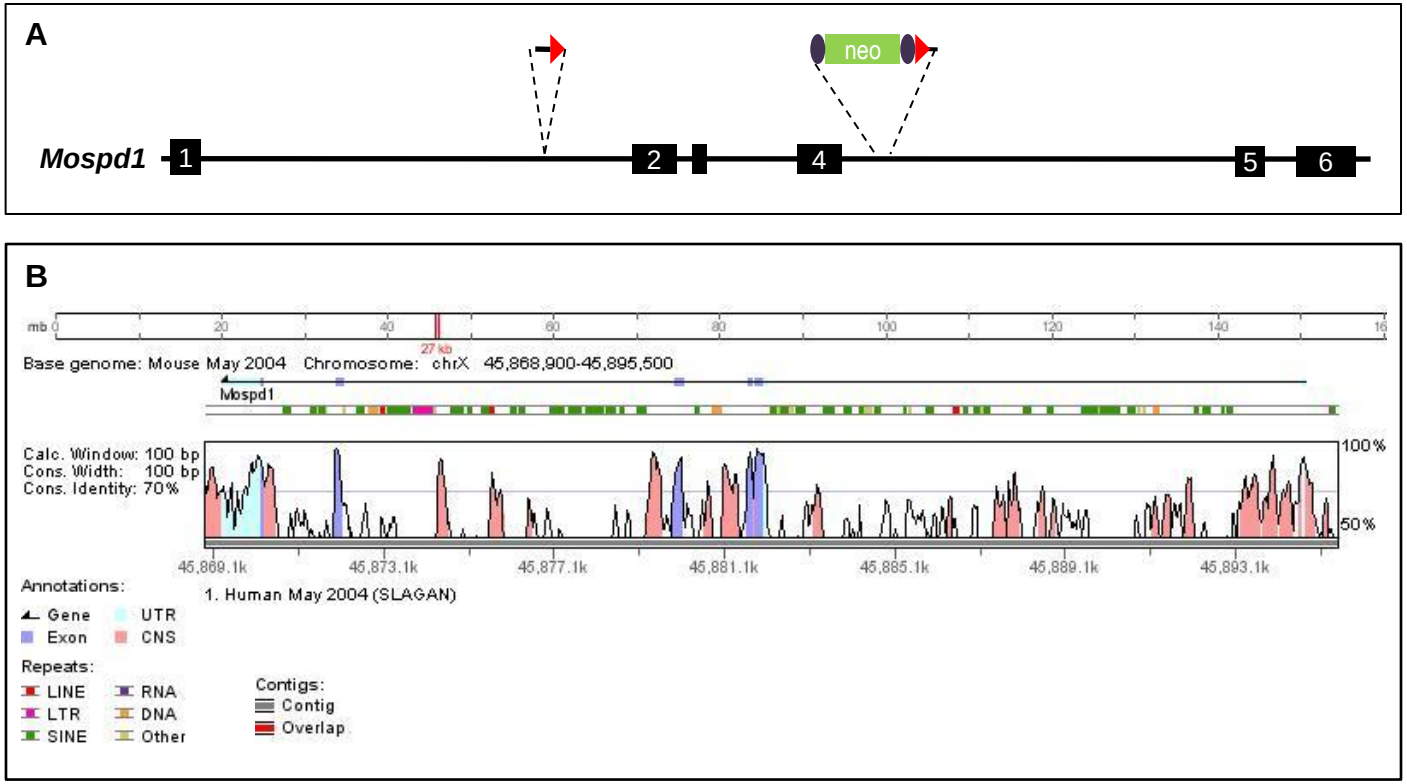
3.2.1.1 Design of the conditional targeting vector

The sequence of *Mospd1* to be targeted, and to be used in the targeting vector, was chosen to include exons 2 to 4 of the gene along with 3 to 4kb of 5'- and 3'-flanking homology arms. Taking account of the genomic structure and protein domain information of *Mospd1,* we chose to place the *loxP* sites in the intron between exons 1 and 2 as well as the intron between exons 4 and 5 (Figure 3.1 A). Being placed in these large (7.7kb and 12.5kb) non-coding intronic regions, the *loxP* sites are not anticipated to interfere with the normal function, nor the expression of *Mospd1*.

To ensure that the *loxP* sites were not positioned in possible conserved non-coding regulatory elements within the introns between exon 1 and 2, and between exons 4 and 5; a genome alignment between mouse and human was performed (Figure 3.1 B). The genomic sequence of *Mospd1* in the NCBI mouse genome build m33 (May 2004) was aligned with the human orthologue from the human reference sequences (genome build 35) published in May 2004. Intronic sequences showing a high level of conservation (more than 70% sequence identity), marked by pink coloured peaks in the alignment, were avoided when placing the *loxP* sites as these conserved sequences could potentially represent regulatory element such as enhancers of *Mospd1* or other genes in the vicinity. The *loxP* integration sites, shown in Figure 3.1, were chosen to minimise the size of the genomic sequence being deleted in the conditional allele whilst including all the important exons - containing both the translational start site (ATG) and MSP domain.

As the start-codon of *Mospd1* is located in the second exon, the recombination of the two *loxP* sites is expected to ablate the entire 5'-coding sequence region of *Mospd1*. The next usable in-frame start codon beyond exon 4 of *Mospd1* is located in exon 5, in the middle of a sequence encoding the first transmembrane domain. Thus, even if an alternative start-codon is utilised, the resulting protein would only contain a short portion of the C-terminal transmembrane domains.

Figure 3.1: Design of the conditional knockout allele of *Mospd1* and genomic sequence alignment of the *Mospd1* gene of mouse and human using VISTA Genome Browser. (A) Schematic of the endogenous *Mospd1* gene locus and proposed placement of the *loxP* recombination sites (one in form of a *FRT-neo-FRT-loxP* cassette) for the generation of the conditional knockout allele of *Mospd1.* the six exons of *Mospd1* are marked as black boxes. In the targeting cassettes containing the *loxP* sites, *loxP* is marked as a red triangle and *FRT* sites are represented as purple ovals. (B) The VISTA graph shows the conservation between *Mospd1* in mouse (genomic build May2004) and *MOSPD1* in human (genomic build May 2004) across the entire genomic sequence of the gene. An arrow signifying the *Mospd1* gene is shown above the graph, with the arrowhead pointing in the direction of the gene. The exons of *Mospd1* are indicated as pale and medium blue blocks. In the VISTA graph, peaks marked in pink indicate non-coding sequences (CNS) of more then 70% sequence identity between mouse and human. More than 70% sequence identity in coding exons is marked by medium blue peaks, and with pale blue peaks in untranslated regions (UTR). The grey bar below the plot indicates that a single human contig was used to generate the alignment with the mouse. The bar immediately above the plot denotes regions of repetitive sequences (listed in the legend under Repeats). The scale bar at the top of the figure shows the size and location of *Mospd1* on mouse chromosome X. Abbreviations: CNS, non-coding sequences, DNA, DNA-transposon; LINE, Long Interspersed Nuclear Elements; LTR, Long Terminal Repeats; neo, neomycin resisance marker; RNA, RNA-retrotransposon; SINE, Short Interspersed Nuclear Elements; UTR, untranslated region.


A
neo
Mospd1
1
2
4
5
6
B
Base genome: Mouse May 2004 Chromosome: chrX 45,868,900-45,895,500
Mospd1
Calc. Window: 100 bp
Cons. Width: 100 bp
Cons. Identity: 70%
100%
50%
45,869.1k
45,873.1k
45,877.1k
45,881.1k
45,885.1k
45,889.1k
45,893.1k
Annotations:
1. Human May 2004 (SLAGAN)
Gene
Exon
UTR
CNS
Repeats:
LINE
LTR
SINE
RNA
DNA
Other
Contigs:
Contig
Overlap

3.2.1.2 Identification of a *Mospd1*-BAC clone

A BAC clone containing the entire genomic sequence of *Mospd1* was chosen using the clone finder tool of the Mouse Genome Resource Database[5]. This BAC clone (GenBank Accession: AC100311), held at the BACPAC facility at the CHORI (Children's Hospital Oakland Research Institute), originated from a BAC library generated from female C57BL/6 DNA, from kidney and brain.

3.2.1.3 Subcloning of *Mospd1*

Although it is possible to use the whole BAC as targeting vector, and engineer the *loxP* sites into *Mospd1*, working with such a large construct (usually around 20kb) has several drawbacks. Due to its size, a BAC has to be handled carefully to keep it from shearing; which, in turn, makes it difficult to examine the vectors integrity after every modification. Additionally, most BACs already contain *loxP* sites which have been used to generate the BAC library. These *loxP* sites would need to be removed before starting to generate the conditional knockout (cko) vector (Liu et al., 2003). Finally, the electroporated ES cell clones would have to be screened and verified for targeting vector integration by FISH (Fluorescent *in situ* Hybridisation); instead of the preferred Southern blot analysis of diagnostic restrictions, which allows more detailed analysis of the site of integration.

An 11kb DNA fragment, containing exons 2 to 4 of *Mospd1* along with flanking homology arms, was directionally subcloned from the BAC into a pNEB-DT vector (see Appendix figure 1 for plasmid map) using *SmaI* and *SacI* restriction sites (Figure 3.2 A). The targeting fragment integrated into the multiple cloning site of the vector, upstream of a negative selection cassette. The negative selection cassette, in this case, consisted of the diphtheria toxin-A gene (DT) driven by a modified MC1 eukaryotic-expression promoter - engineered from the promoter sequences of Herpes Simplex Virus thymidine kinase (Pinto et al., 2000; Thomas and Capecchi, 1987; Yagi et al., 1990). This cassette also contains a polyadenylation signal (Yanagawa et al., 1999).

[5] www.ncbi.nih.gov/genome/guide/mouse/

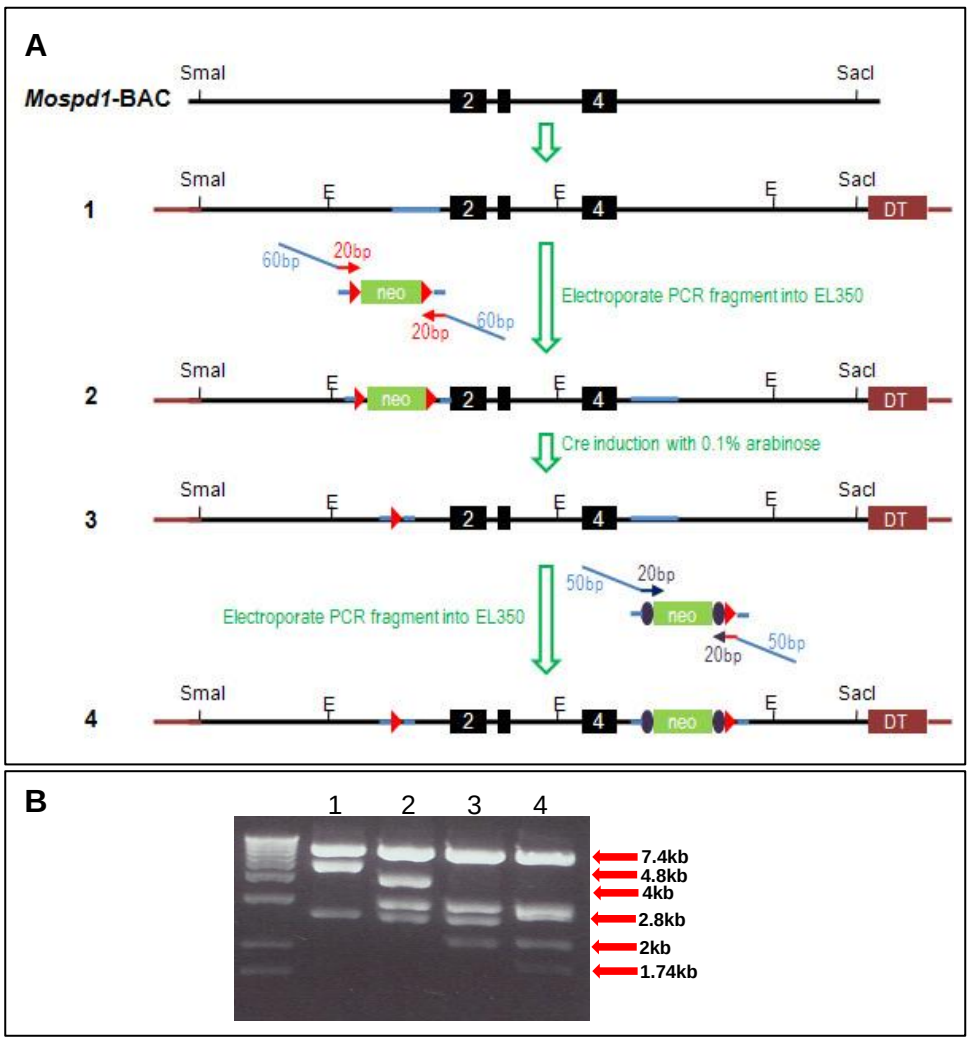


Figure 3.2: Recombineering strategy for *Mospd1* targeting vector construction. (A) Schematic of recombineering strategy for generating a conditional targeting vector of *Mospd1* and (B) diagnostic *EcoRI* digests of constructs generated in this procedure. An 11kb targeting fragment of *Mospd1* was subcloned from the *Mospd1*-BAC into pNEB-DT using *SmaI* and *SacI* restriction sites (construct 1 and Lane1 in *EcoRI* digest gel). The exons of Mospd1 are marked by black boxes. The *Mospd1* targeting vector was then electroporated into EL350 cells for further manipulation. The first *loxP* site (*loxP* marked as red triangle) was introduced into the vector by recombineering a PCR amplified *loxP-neo-loxP* cassette into the sequence upstream of exon 2 of *Mospd1* in the vector (A, construct 2 ;B, lane 2). Recombineering-specificity to this region was achieved through the 60bp sequence homology of the PCR with sequences upstream of exon 2. Homologies for targeting of the selection cassettes are marked as blue lines in the constructs. Subsequently, the neo selection cassette was removed by inducing the EL350 cells to express Cre recombinase by addition of 0.1% arabinose (A, construct 3; B, lane 3). The second *loxP* site in the form of a *FRT-neo-FRT-loxP* cassette (*FRT* site marked with purple oval) was introduced by recombination downstream of *Mospd1* exon 4 in the same way as the first targeting cassette (A, construct 4 and B, lane 4). A 1kb plus ladder from Invitrogen was used for gel electrophoresis (B). Abbreviations: DT, diphtheria toxin; E, *EcoRI*; neo, neomycin resistance marker.

Diagnostic digests were used to confirm the correct subcloning of the 11kb targeting fragment. The presence of this additional sequence in the pNEB-DT parental plasmid, which itself is not cut by *EcoRI* (data not shown), resulted in the release of 3 fragments (7.4kb, 4.8kb and 2.6kb), when digested with *EcoRI* (Figure 3.2 B (lane 1)).

In preparation for further modifications, the targeting vector *Mospd1*-pNEB-DT was electroporated into recombination efficient EL350 *E.coli* cells, and maintained under ampicillin selection.

3.2.1.4 Introduction of the first *loxP* site

The next step, towards generating the *Mospd1* targeting vector, involved the introduction of a *loxP* site into the intron between exons 1 and 2 of the subcloned DNA (Figure 3.2 A). To this end, a f*lox*ed neomycin resistance (neo) cassette was integrated *via* homologous recombination into the vector DNA. The cassette was amplified from plasmid PL452 (Appendix figure 2) with 80mer primers (Appendix table 1). To achieve homologous recombination of the cassette into the subcloned plasmid DNA, the flanking 60bp of the PCR primers consisted of sequence homologous to the target intron of *Mospd1*, where the cassette was to be placed; and 20bp of the primers was homologous to the f*lox*ed *Neo* cassette, to be amplified.

The purified PCR product was electroporated into EL350 cells containing the *Mospd1*-pNEB-DT targeting vector. The EL350 cells were grown at 42ºC for 15 minutes, prior to electroporation, to induce the expression of the *Red* recombination genes. EL350 cells carrying the f*lox*ed cassette in the *Mospd1*-pNEB-DT targeting vector were selected with kanamycin.

Sequencing across the site of integration, using flanking primers, as well as diagnostic digests confirmed the presence of the f*lox*ed neomycin resistance cassette. The presence of the approximately 2kb f*lox*ed neo cassette resulted in the introduction of an additional *EcoRI* site into the targeting vector. *EcoRI* digestion of this new construct produced two separate fragments of 4kb and 2.8kb (Figure 3.2 B, lane 2), instead of the single 4.8kb fragment seen in the previous construct (Fig. 3.2 B, lane 1).

3.2.1.5 Removal of the *neo* resistance marker

Once the correct targeting vector construct, containing the f*lox*ed neo cassette, was identified using kanamycin selection, the neomycin resistance gene was removed to create a single *loxP* site. This was achieved by electroporation of the targeting vector, containing the f*lox*ed neo cassette, into EL350 cells which had been grown for 1 hour in arabinose-supplemented medium. This prior incubation led to the expression of Cre recombinase from an arabinose-inducible promoter. The high efficiency of Cre-induced recombination in these cells (Liu et al., 2003) has facilitated the recombination of the two *loxP* sites to allow excision of the resistance marker in all of the electroporated cells. When the electroporated cells were spread on antibiotic selection plates, colonies were only observed on ampicillin; but not on kanamycin. In addition, colonies picked from ampicillin plates were found to be sensitive to kanamycin due to the absence of the deleted neomycin resistance marker (data not shown).

The excision of the neo resistance marker in these clones was verified by sequencing across the remaining *loxP* site, as well as by diagnostic digestion. The removal of the 2kb *neo* sequence, within the 4kb *EcoRI* fragment (seen in Figure 3.2B, lane 2), resulted in a 2kb *EcoRI* restriction fragment (Figure 3.2 B, lane 3).

3.2.1.6 Introduction of the second *loxP* site

In order to introduce the second *loxP* site 3' of *Mospd1* exon 4 in the targeting vector, a cassette containing a neo resistance gene flanked by two *FRT* sites, and followed by a single *loxP* site, was used.

This cassette was used, instead of the f*lox*ed neo cassette, to ensure that the subsequent removal of this neo resistance marker in ES cells would not result in a knockout allele. As the neo resistance marker is required to select for the integration of the targeting vector into ES cells, it cannot be removed until the correctly targeted ES cell clone is confirmed. If a f*lox*ed neo cassette was used, Cre recombination of this conditional allele would result in 1 of 3 different excision products due to the presence of 3 *loxP* sites in the allele (see Chapter 1 figure 1.9 C). However, using a cassette containing neo^R flanked by *FRT* sites - coupled with transient expression of flp recombinase, which acts on *FRT* sites, will lead to the removal of the neomycin resistance marker without affecting the rest of the *Mospd1* targeted allele.

The *FRT-neo-FRT-loxP* cassette was PCR amplified from plasmid PL451 (Appendix figure 3) using 70mer primers containing: a 50bp flanking sequence specific to the target sequence of the cassette in intron 4-5 of *Mospd1*, and 20bp of sequence specific to the *FRT-neo-FRT-loxP* cassette itself (primer sequences in Appendix table 1). The purified PCR product was electroporated into red-recombination induced EL350 cells containing the *Mospd1* targeting vector (Figure.3.2A). Cells having integrated the cassette were selected with kanamycin.

The presence of the *FRT-neo-FRT-loxP* cassette and the integrity of the targeting vector were confirmed by sequencing and diagnostic digests. With the introduction of the *FRT-neo-FRT-loxP* cassette, a new *EcoRI* site is added to the restriction pattern; thereby, decreasing the size of the 2.6kb fragment to 1.74kb, and generating a new fragment of 2.8kb from the remaining 0.9kb plus the 1.9kb of the *FRT-neo-FRT-loxP* cassette (Figure 3.2 B, lane 4). Once the sequence integrity of the *loxP* sites has been confirmed, the targeting vector was ready to be linearized and electroporated into ES cells.

3.2.1.7 Linearization and electroporation of the targeting vector into ES cells and selection of *neo* resistant clones

The 17kb targeting vector, containing exons 2 to 4 of *Mospd1* flanked by *loxP* sites and, both, a 4kb 5' homology and a 3kb 3' homology; was linearized using *PmeI*, a unique restriction site present in the multiple cloning site of pNEB-DT.

Two different ES cell lines (E14 and CGR8) were electroporated with the linearized *Mospd1* targeting vector. The electroporated cells were subsequently selected for their resistance to G418, conferred by the neomycin resistance marker in the *FRT-neo-FRT-loxP* cassette. In addition, non-homologous recombination events were selected against by the diphtheria toxin (DT) cassette present in the targeting construct.

After 6 to 8 days under selection, ES cell clones were picked into 96-well tissue culture plates, and subsequently split in triplicate. One master plate was frozen whilst DNA was extracted from the other two.

3.2.1.8 Design of the Southern blot strategy to detect *Mospd1* targeted ES cell clones

In order to allow identification of ES cell clones targeted at the *Mospd1* locus, a Southern blot strategy was designed, with help of the DNA Star software, and tested using wild type genomic DNA from CGR8 ES cells. Both, the 5' and 3' probe used were designed to label genomic sequences outwith the *Mospd1* sequence contained in the targeting vector (Figure 3.3 A).

The Southern blot probes were hybridised to blots of wild type CGR8 DNA digested with various restriction enzymes, which generated different sized restriction fragments in *Mospd1* wild type, targeted or null allele (Figure 3.3B). Analysis of these test Southern blots showed that the 5' probe bound well to the genomic target sequence and exhibited a good level of radioactive signal. However, the 3' probe showed a less distinctive signal.

ES cell clones containing the targeting vector integration at the *Mospd1* locus could be distinguished from wild type cell by DNA digestion, with either *Spe1* or *Pst1* restriction enzyme, and labelling with the 5' probe (Figure 3.3 A). 5' probe labelling of *Spe1* digested DNA resulted in a distinct 14.7kb signal (Figure 3.3 B), which shifted to 7kb in *Mospd1* targeted ES cell clones due to the presence of a *Spe1* site in the *loxP* site proceeding exon 2 of *Mospd1* in the targeting vector . The shift in fragment sizes using *Pst1* restriction was smaller. In this case, the wild type fragment of about 9kb was shifted to an 8kb fragment due to the insertion of a new *Pst1* site in the upstream *loxP* sequence of the targeting vector (Figure 3.3 A). As the ES cell lines (CGR8 and E14) used to generate the targeted allele are male, and mouse *Mospd1* is an X-linked gene, a single band is expected in the wild type Southern blot. This signal should shift in size if the allele is correctly targeted.

Generation of a null allele of *Mospd1* can be detected using 5' probe labelling of a *Bcl1* restriction digest. The loss of the *Bcl1* restriction site in exon 3 of *Mospd1,* through the *Cre* induced recombination of the conditional allele, led to a shift of the 7.7kb targeted fragment to 17.7kb in the null allele (3.3 A and B).

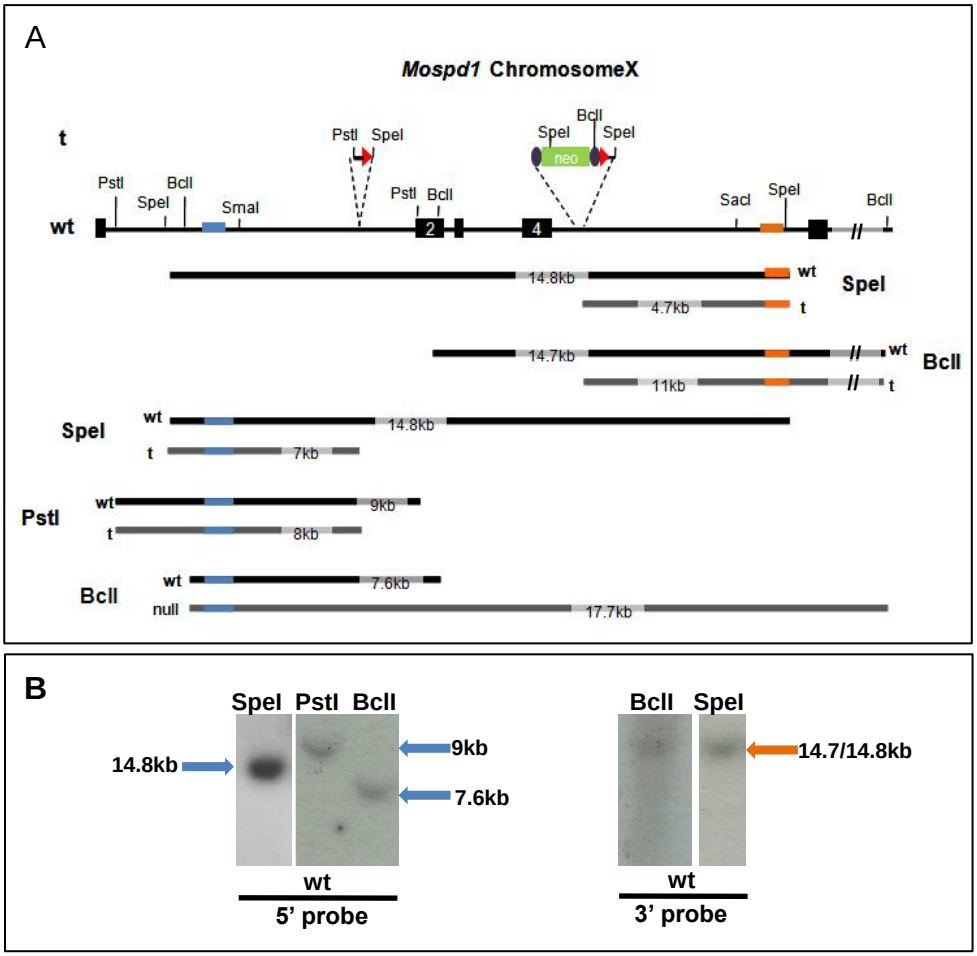


Figure 3.3: Design and testing of Southern blot screening strategy for *Mospd1* targeted clones. (A) Design of 5' and 3' probe hybridisation to different restriction enzyme fragments of genomic DNA in order to distinguish between wild type, targeted and null alleles in ES cells. The upper panel in (A) shows the wild type allele of *Mospd1* as well as the integration sites of the single *loxP* site (marked as red triangle) and *FRT-neo-loxP* cassette (*FRT* site marked as purple oval) to generate the targeted allele. The lower panel of (A) shows the anticipated restriction fragments detected with 5' probe (blue line) or 3' probe (orange line) labelling of different restriction digests of genomic DNA. The solid black lines mark restriction fragments anticipated in wild type alleles (wt) whilst the grey lines mark the restriction fragments anticipated for targeted (t) or null alleles. (B) Southern blot verification of anticipated wild type restriction fragment sizes using 5' and 3' probe labelling of genomic DNA isolated from wild type (wt) CGR8 ES cells. Abbreviations: neo, neomycin resistance marker; t, targeted allele; wt, wild type allele or wild type ES cell genomic DNA.

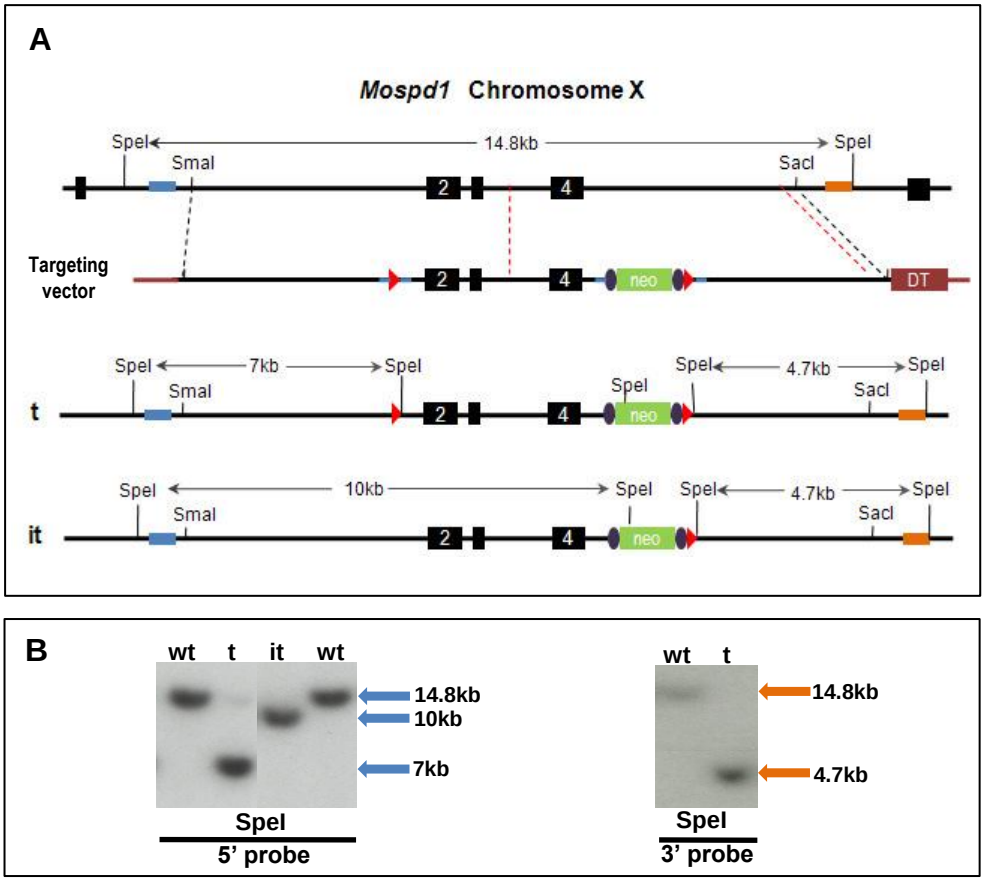


Figure 3.4: Southern blot analysis of *Mospd1* targeted ES cell clones.
(A) Schematic showing a map of the *Mospd1* genomic locus (wild type allele, wt), the targeting vector and the *Mospd1* targeted allele (t) as well as a targeted allele generated from incomplete target vector integration (it). The exons of *Mospd1* are represented by black boxes. The Southern blot probes used are marked by blue boxes for the 5' probe and orange boxes for the 3' probe. The targeting vector construct includes the relative position of the *loxP* site (red triangles), the *FRT–neo-FRT-loxP* cassette (*FRT,* purple ovals) and the diphtheria toxin (DT) cassette. Whilst the black dashed lines represent the crossover event between the wild type allele and targeting vector leading to integration of the both *loxP* and *FRT–neo-FRT-loxP* cassette in the targeted allele (t), the red dashed line represents the cross-over event leading to incomplete vector integration and lack of *loxP* site in the targeted allele (it). (B) Southern blot analysis of ES cell genomic DNA to screen for ES cells carrying a correctly targeted allele of *Mospd1*. Abbreviations: DT, diphtheria toxin cassette; it, clone containing an incomplete targeting vector integration into the *Mospd1* locus; neo, neomycin resistance marker; t, targeted allele or targeted ES cell clone, wt, wild type allele or ES clone.

3.2.1.9 Southern blot screening of *Mospd1* targeted ES cell clones

Southern blot analysis of the G418 resistant clones, using a 5' probe on *Spe1* digested genomic DNA, was used to identify correctly targeted clones (Figure 3.4 A). Southern blot screening of G418 resistant ES cell clones identified 6 correctly targeted clones out of 351 in CGR8 cells, and 4 correctly targeted clones out of 179 in E14 ES cells. These numbers correspond to a targeting frequency of 1.7% in CGR8, and 2.2% in E14 cells. Integration of the targeting vector at the *Mospd1* locus, in these clones, resulted in the expected shift in restriction fragment sizes from 14.7kb (wild type) to 7kb (targeted allele) (Figure 3.4 B).

Among the G418 resistant CGR8 clones, 18 showed a *SpeI* restriction fragment size of about 10kb on the Southern blot (Figure 3.4 B, lane (it)). Both, the single *loxP* site upstream of *Mospd1* exon 2, and the *FRT-neo-FRT-loxP* cassette in the intron between exons 4 and 5, contain *Spe1* restriction sites. Instead of cutting at the *loxP* site in the intron between exon 1 and 2 of *Mospd1* resulting in a restriction fragment of 7kb, the next *Spe1* digestion site 3kb downstream in the *FRT-neo-FRT-loxP* cassette was utilised. This suggested that the upstream *loxP* site of the targeting vector had not integrated into the endogenous *Mospd1* locus. The vector integration, in these cases, seemed to have occurred *via* one homologous recombination event in the 3' homology arm, and a second recombination event downstream of the first *loxP* site instead of within the 5' homology arm. This type of crossover event was expected, as the sequences (including exons 2 to 4) between the two *loxP* cassettes of the vector were also homologous to the wild type *Mospd1* allele.

3.2.1.10 Expansion and verification of targeted clones

The *Mospd1* targeted ES cell (CGR8 and E14) clones, identified from the Southern blot screen, were thawed from the master 96-well plate and expanded. Genomic DNA was extracted for further Southern blot verification using a 5' and 3' probe as well as a Neo^R specific sequence probe (neo probe). Both, 5' and 3' probes, in conjunction with *SpeI* restriction, confirmed that in five out of six of the CGR8 derived clones the targeting vector had integrated at the *Mospd1* locus. This analysis also confirmed that, both, the upstream *loxP* site and the *FRT-neo-FRT-loxP* cassette were present in the targeted allele (Figure. 3.4 A and B). E14 targeted clones have been expanded, from the master plate, for DNA extraction. Southern blot

confirmation of the four *Mospd1* targeted clones in E14 cell is currently underway (performed by Madina Kara).

Using the neo (neomycin cassette specific) Southern blot probe, genomic ES cell sequences can be screened for unwanted multiple targeting vector integrations. In all 5 targeted clones of *Mospd1*, derived from CGR8 cells, a single *EcoRV* restriction fragment was labelled, indicating the presence of only one vector integration (data not shown).

3.2.1.11 Generation of a *Mospd1* null allele in ES cells

In order to analyse whether the loss of *Mospd1* had any effect on the physiology of ES cells themselves, a knockout allele of the gene was produced.

This null allele was generated by transient transfection of two of the targeted *Mospd1* clones with a *Cre* expression vector (pCAGGS-Cre-IRESpuro) (see Appendix figure 4 for vector map). The Cre recombinase, expressed in these cells, drove the recombination of the *loxP* sites in the conditional allele, resulting in an excision of exon 2 to 4 of *Mospd1* and the production of a loss-of-function knockout allele (Figure 3.5 A).

Clones containing the knockout allele were identified by Southern blot screening, using the *Mospd1* 3' probe to label *HindIII* restriction fragments in digested genomic DNA. A shift of fragment sizes from the 6.7kb of the targeted allele to a larger fragment of 8.3kb, due to the loss of a *HindIII* site present in *Mospd1* exon3, confirmed the generation of the null allele (Figure 3.5 B).

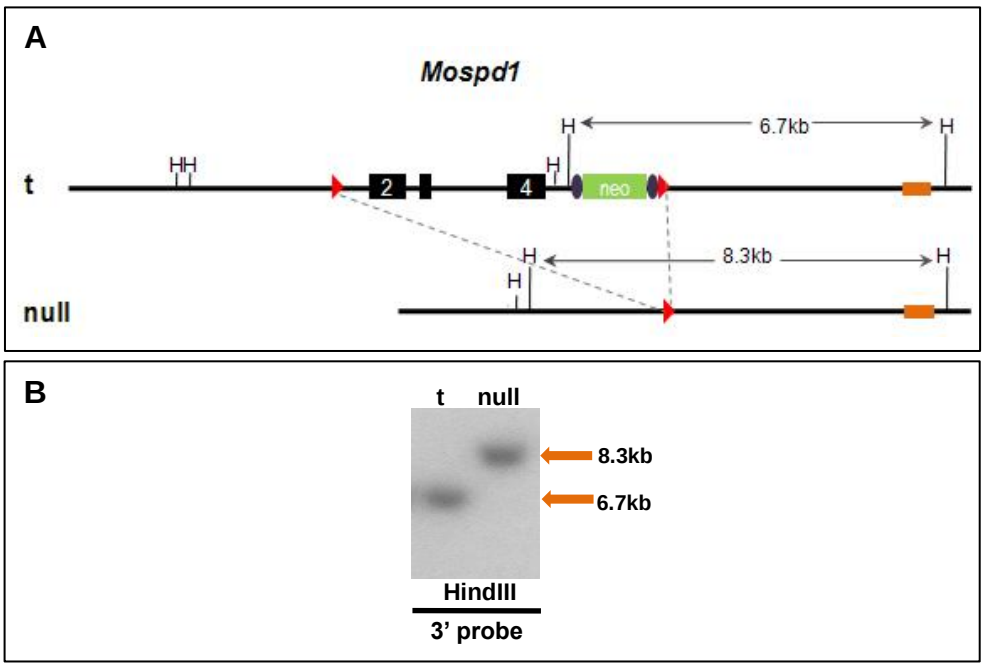


Figure 3.5: Southern blot analysis of *Mospd1* null ES cell clones.
(A) Schematic showing a map of the *Mospd1* genomic locus (t) and the *Mospd1* null allele after Cre mediated recombination of the *loxP* sites (marked by red triangles), which mediates deletion of the *FRT-neo-FRT-loxP* cassette (*FRT* marked by purple oval) as well as exons 2 to 4 of *Mospd1*. The exons of *Mospd1* are represented by black boxes. The 3' Southern blot probe used to distinguish between *HindIII* (H) restriction fragments of the targeted and the null allele of *Mospd1* is marked by orange boxes. (B) Southern blot analysis of ES cell genomic DNA to screen for ES cells carrying a *loxP* site-specific recombination at the *Mospd1* locus. Abbreviations: H, *HindIII*; neo, neomycin resistance marker; null, *loxP* -recombined allele or ES cell clone carrying a recombined null allele; t, targeted allele or targeted ES cell clone.

3.2.2 Generation of *Mospd3* targeting construct and targeted allele

3.2.2.1 Design of the conditional targeting vector

In order to generate a conditional knockout allele of *Mospd3*, the targeting vector was designed to include the start-codon in exon 2 and the MSP domain encoded in exons 3 and 4 (Figure 3.6 A).

Unlike for *Mospd1*, it was not possible to place the 5' *loxP* site within the intron between exons 1 and 2 of *Mospd3*. Due to the short sequence length of this intron (198bp), the presence of a *loxP* site and its flanking restriction site sequences (100bp) might interfere with correct splicing. A genome sequence alignment between mouse and human also revealed that a large region of this intron contained sequences which appear to be conserved between the two species (Figure 3.6 B). In fact, the alignment indicates the presence of two peaks of sequence conservation in the vicinity of exon 1, which may indicate promoter or enhancer elements of *Mospd3*. In view of these findings, we opted to place the 5' *loxP* site upstream of exon 1 (as shown in Figure 3.6 A). The 3' *loxP* site was to be placed in the intron between exons 4 and 5, which according to the genome alignment was devoid of any regions of significant sequence conservation.

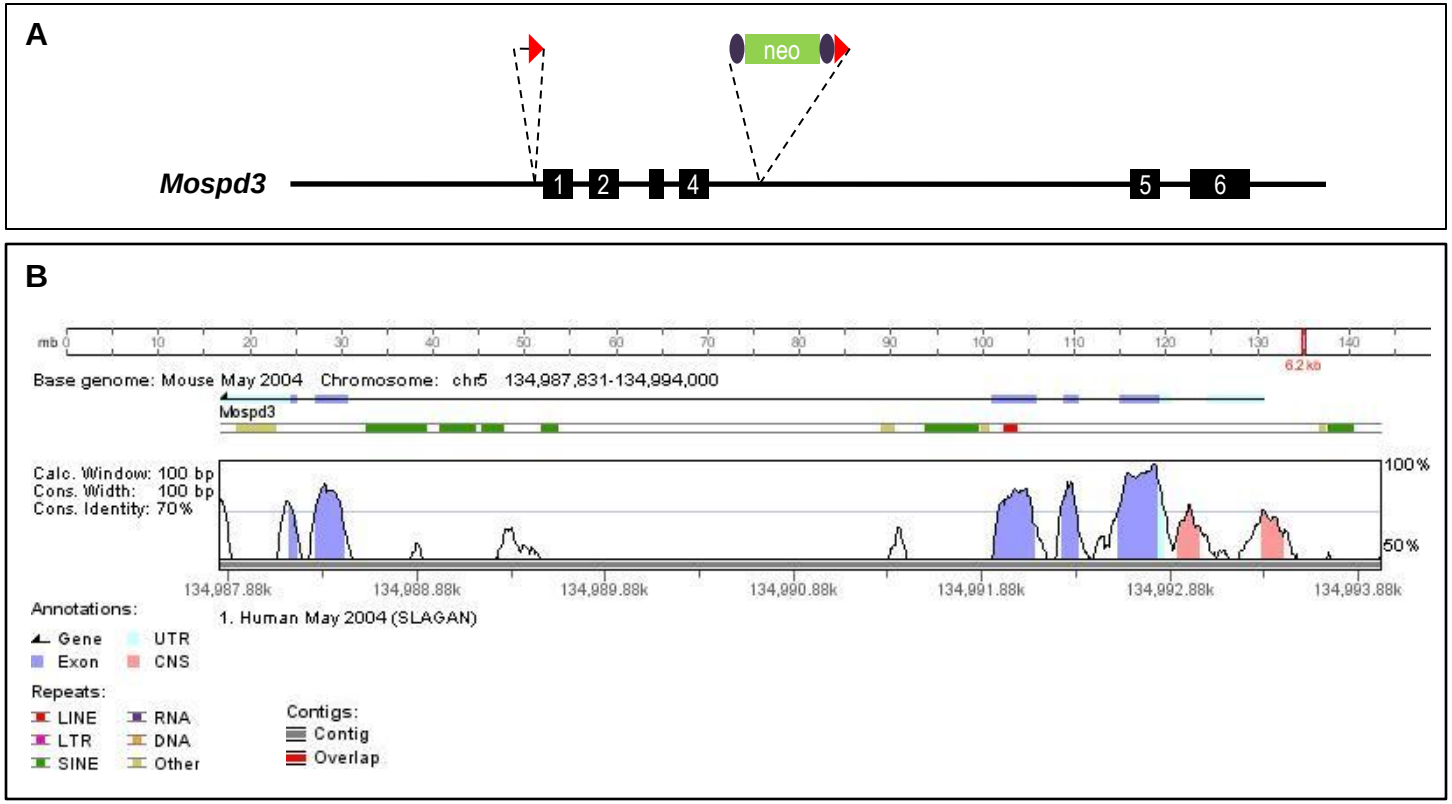
Cre induced recombination of the *loxP* sites, in this conditional allele, is anticipated to lead to a complete loss of the Mospd3 transcript, as not only the start codon but also the putative promoter of the gene will be removed.

3.2.2.2 Identification of a *Mospd3*- BAC clone

Using the aforementioned Mouse Genome Resource [6], a BAC clone (GenBank Accession: BH063983), containing the entire genomic sequence of *Mospd3*, was chosen to subclone the targeting sequence. This BAC clone was part of a male C57BL/6 library, generated from spleen and brain DNA.

[6] www.ncbi.nih.gov/genome/guide/mouse/

Figure 3.6: Design of the conditional knockout allele of *Mospd3* and genomic sequence alignment of the *Mospd3* gene of mouse and human using VISTA Genome Browser. (A) Schematic of the endogenous *Mospd3* gene locus and proposed placement of the *loxP* recombination sites (one in form of a *FRT-neo-FRT-loxP* cassette) for the generation of the conditional knockout allele of *Mospd3.* The six exons of *Mospd3* are marked as black boxes. In the targeting cassettes containing the *loxP* sites, *loxP* is marked as a red triangle and *FRT* sites are represented as purple ovals. (B)The VISTA graph shows the conservation between *Mospd3* in mouse (genomic build May2004) and *MOSPD3* in human (genomic build May 2004) across the entire genomic sequence of the gene. The VISTA graph also contains alignments of approximately 700bp upstream of exon 1 of *Mospd3*. An arrow signifying the *Mospd3* gene is shown above the graph, with the arrowhead pointing in the direction of the gene. The exons of *Mospd3* are indicated as pale and medium blue blocks. In the VISTA graph, peaks marked in pink indicate non-coding sequences (CNS) of more then 70% sequence identity between mouse and human. More than 70% sequence identity in coding exons is marked by medium blue peaks, and with pale blue peaks in untranslated regions (UTR). The grey bar below the plot indicates that a single human contig was used to generate the alignment with the mouse. The bar immediately above the plot denotes regions of repetitive sequences (listed in the legend under Repeats). The scale bar at the top of the figure shows the size and location of *Mospd3* on mouse chromosome 5. Abbreviations: CNS, non-coding sequences, DNA, DNA-transposon; LINE, Long Interspersed Nuclear Elements; LTR, Long Terminal Repeats; neo, neomycin resistance marker; RNA, RNA-retrotransposon; SINE, Short Interspersed Nuclear Elements; UTR, untranslated region.


A
neo
Mospd3
1
2
4
5
6
B
mb 0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
6.2 kb
Base genome: Mouse May 2004 Chromosome: chr5 134,987,831-134,994,000
Mospd3
Calc. Window: 100 bp
Cons. Width: 100 bp
Cons. Identity: 70%
100%
50%
134,987.88k 134,988.88k 134,989.88k 134,990.88k 134,991.88k 134,992.88k 134,993.88k
Annotations:
1. Human May 2004 (SLAGAN)
Gene
Exon
UTR
CNS
Repeats:
LINE
LTR
SINE
RNA
DNA
Other
Contigs:
Contig
Overlap

3.2.2.3 Subcloning of *Mospd3* by Gap repair

The initial strategy for subcloning a sufficiently long *Mospd3* sequence, containing the exons to be flanked by *loxP* sites as well as a minimum of 3kb homology arms, involved the use of *XbaI* restriction and ligation into the pNEB-DT plasmid. The chosen *Xba1* fragment contained the entire genomic sequence of *Mospd3*. However, this subcloning method proved unsuccessful as the pNEB-DT plasmid appeared to contain an additional *XbaI* site. A digest with this enzyme revealed two restriction sites: one, in the multiple cloning site of the vector; and the other, in the diphtheria toxin (DT) cassette. Digestion of the DT cassette, a negative selection cassette for targeting vector integration into ES cells, would have caused problems at a later stage when selecting for homologous vector integrations into ES cells.

As an alternative to using specific restriction sites for the subcloning step, a gap repair approach in *E.coli* cells was utilised (Liu et al., 2003). This homologous recombination approach is based on the repair of a double strand gap in the linearized pNEB-DT plasmid by DNA synthesis with a homologous DNA as a template.

Sequences homologous to regions flanking the 10.6kb *Mospd3* targeting DNA, in the BAC, were PCR amplified (Figure 3.7 A). These, approximately 200bp, sequences were directionally subcloned into pNEB-DT using restriction sites engineered into the PCR primers. The resulting retrieval vector, linearized with *BamH1*, was electroporated into recombination efficient EL350 cells containing the *Mospd3* BAC. The double-strand break of the linearized vector was repaired by homologous recombination with *Mospd3* sequences in the BAC. This process allowed the genomic sequences of *Mospd3* to be integrated into the multi-copy pNEB-DT plasmid.

The electroporated clones were selected using ampicillin. Most of the clones obtained and analysed, by sequencing and diagnostic digests, contained the 10.6kb *Mospd3* targeting sequence within the pNEB-DT vector backbone (Figure 3.7 A, construct 1). The *EcoR1* digest in (Figure 3.7 B) confirmed the presence of the correct fragment, as the pNEB-DT vector, which had no *EcoRI* site, could now be cut by the enzyme; and the restriction fragments of 9kb, 3.8kb 1.7kb and 0.46kb were

those expected from the presence of the 11.5kb *Mospd3* fragment (Figure 3.7 B, lane 1).

3.2.2.4 Introduction of the first *loxP* site

The f*lox*ed neo cassette containing the upstream *loxP* site was PCR amplified from plasmid PL452 (Appendix figure 2) in a similar way to that described for *Mospd1*. The primers, which were used, contained 60bp of sequence homologous to the genomic region preceding exon 1 of *Mospd3* and 20bp of sequence homologous to the neo targeting cassette. The PCR amplified cassette was electroporated into recombination efficient EL350 cell containing the *Mospd3*-pNEB-DT targeting vector.

Kanamycin resistant clones, having integrated the cassette, were analysed by diagnostic restriction digestion and sequencing across the integration site. From the *EcoRI* restriction digest, it was observed that not all the multi-copy *Mospd3-pNEB-DT* plasmids, which have integrated into a single EL350 cell, contained the f*lox*ed neo cassette (data not shown). This can sometimes occurs because one single plasmid containing the neo resistance marker is sufficient in an electroporated EL350 cell to confer resistance to kanamycin. Re-electroporation of 1ng of the vector DNA, isolated from the original electroporated clones, allowed the retrieval of colonies in which all plasmids contained the f*lox*ed neo cassette.

3.2.2.5 Removal of the *neo* resistance marker

The neo resistance marker was removed by electroporation of the *Mospd3* targeting clone into arabinose-induced *Cre* expressing EL350 cells. Through the action of Cre, the two *loxP* sites recombine, and the neo resistance marker was excised in the process; rendering the resulting bacterial colonies kanamycin sensitive (Figure 3.7 A, construct 3).

The presence of the single 100bp *loxP* site, which does not change the *EcoRI* digestion pattern of the *Mospd3* targeting vector significantly (Figure 3.7 B, lane 3), was confirmed by sequencing.

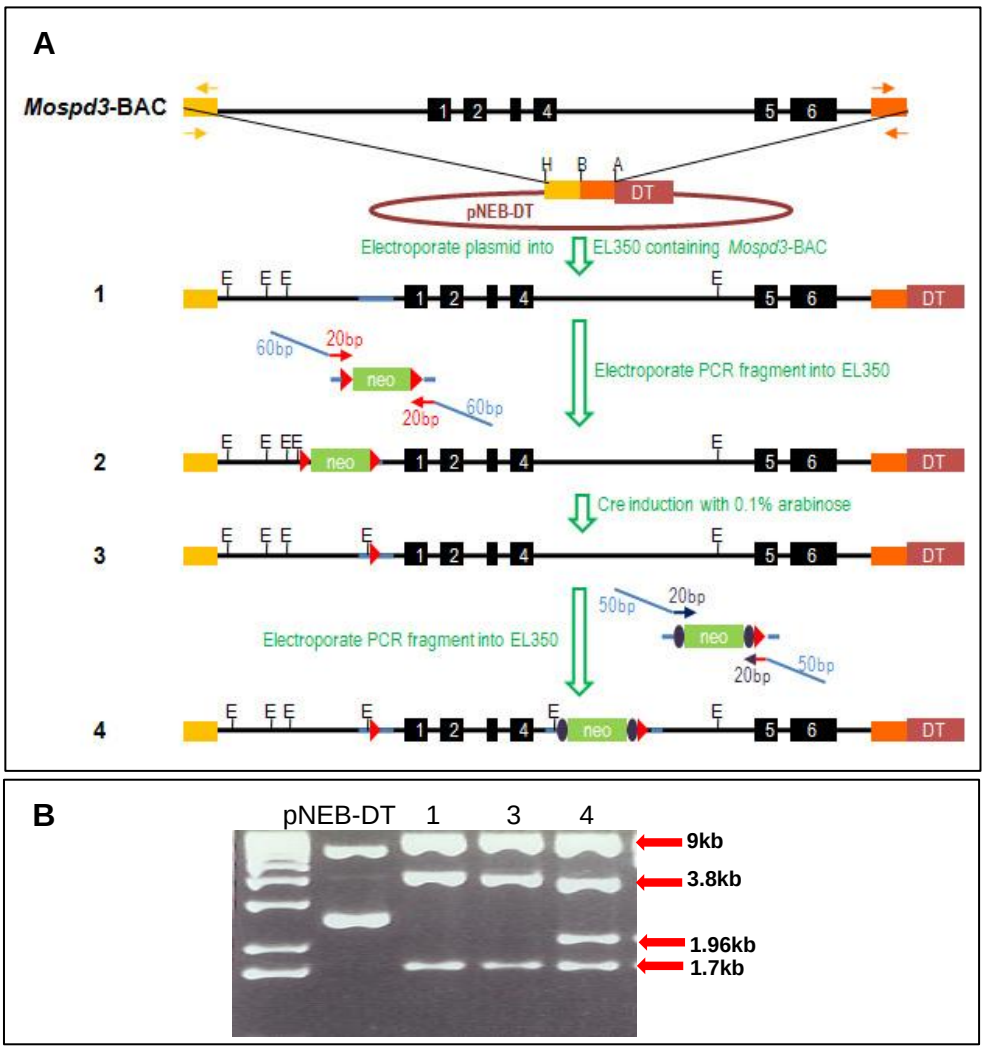


Figure 3.7: Recombineering strategy for *Mospd3* targeting vector construction. (A) Schematic of the recombineering strategy used to generate a conditional *Mospd3* targeting vector and (B) *EcoRI* diagnostic digest of successive constructs generated in the procedure. The 10.6kb targeting fragment containing *Mospd3* exons 1 to 6 (exons marked by black boxes) was subcloned into the pNEB-DT vector (A, construct 1 and B, lane 1) using gap repair between two approximately 200bp homology regions (yellow and orange box) flanking the 10.6kb fragment. These homologous regions were PCR amplified from the *Mospd3*-BAC and subcloned into pNEB-DT using the *HindIII*, *BamHI* and *AvaI* restriction sites engineered into the PCR primers (yellow and orange arrows). The *BamHI* linearised pNEB-DT retrieval vector was electroporated into EL350 cells, which already contained the *Mospd3*-BAC, leading to subcloning of the 10.6kb fragment into the pNEB-DT vector. The PCR amplified *loxP-neo-loxP* targeting cassette (*loxP* sites marked as red triangles) containing 60bp flanking sequence homologous to sequences upstream of *Mospd3* exon1 (homologies marked as blue lines) has integrated into the targeting construct (A, construct 2). Subsequently, the *neo* casstte was removed by arabinose-induced Cre expression (A, construct 3 and B, lane 3) leaving a single *loxP* site upstream of exon 1. The second *loxP* site downstream of *Mospd3* exon 4 was integrated in the form of a PCR amplified *FRT-neo-FRT-loxP* cassette (*FRT* site maked with purple oval) (A, construct 4 and B, lane 4) in the same way as the previous *loxP-neo-loxP* cassette. A 1kb plus ladder from Invitrogen was used for gel electrophoresis (B). Abbreviations: DT, diphtheria toxin; E, *EcoRI; neo,* neomycin resistance marker.

3.2.2.6 Introduction of the second *loxP* site

The *FRT-neo-FRT-loxP* cassette, containing the second *loxP* site, was PCR amplified from vector PL451 (Appendix figure 3). The 70mer primers, used for this amplification, included a 50bp homology to the sequence downstream of *Mospd3* exon 4 into which the *loxP* site was to be integrated, as well as a 20bp sequence homologous to the *FRT-neo-FRT-loxP* cassette (primer sequence in Appendix table 1). The recombineering procedure used was similar to that described for the *Mospd1* targeting vector. Once again, the targeting vector had to be re-electroporated into EL350 cells to obtain pure colonies, which contained only targeting vectors with the *FRT-neo-FRT-loxP* cassette (Figure 3.7 A, construct 4).

The addition of this new cassette resulted in the introduction of a new *EcoRI* site into the targeting vector; thereby, shortening the 3.70kb fragment to 3.54kb, and adding a new fragment of 1.96kb (Figure 3.7 B, lane 4).

3.2.2.7 Linearization and electroporation of the targeting vector into ES cells and selection of *neo* resistant clones

When the 200bp homologies for gap retrieval subcloning of the *Mospd3* targeting vector sequence were designed, the rare restriction site *PmeI* was lost from the pNEB-DT plasmid. This meant that, in order to linearize the targeting construct, a unique restriction site (*AflII*) present in the 5' homology arm of *Mospd3* had to be utilised. Unfortunately, this approach resulted in the loss of about 2kb of upstream homology of the vector.

CGR8 and E14 ES cells were electroporated with the linearized DNA of the truncated *Mospd3* targeting vector. The electroporated ES cells were selected using G418. After 6-8 days, clones were picked and further processed in the same way as previously described for *Mospd1*.

3.2.2.8 Design of the Southern blot strategy to detect *Mospd3* targeted ES cell clones

A Southern blot strategy to distinguish *Mospd3* targeted and null allele from the wild type, by using diagnostic restriction digests, was designed using DNA star software. In order to identify these different alleles, 5' and 3' probes were designed to sequences flanking the site of the targeting vector integration, and tested on

restriction digests of wild type CGR8 genomic DNA (Figure 3.8 A). Both probes were found to label the hybridised DNA well, resulting in a distinct band of the expected size on the Southern blot of *SpeI* and *KpnI* digests (Figure 3.8 B). *XhoI* restriction digestion, chosen specifically to identify the *Mospd3*-null allele, appears to be unsuitable for Southern blot identification of targeted ES cell clones as the 3' probe was found to label two different wild type fragments (Figure 3.8). The presence of two restriction fragments is likely due to incomplete DNA digestion with *XhoI* as this enzyme can be affected by CpG base pair methylation in the restriction site sequence.

3.2.2.9 Southern blot screening of *Mospd3* targeted ES cell clones

The 3' probe was used to screen for *SpeI* restriction products in the genomic DNA of G418 resistant clones. The presence of a targeting vector integration at the *Mospd3* locus led to a shift in the restriction fragment size of one of the wild type *Mospd3* alleles. The *Spe1* fragment in the targeted allele was 7.8kb in size, compared to 12.7kb in the wild type allele (Figure 3.9 A and B). Out of 320 screened clones for CGR8 ES cells, 5 were found to contain, both, a wild type as well as the targeted restriction fragment. This corresponds to a targeting frequency of 1.5% in CGR8 cells. Amongst the G418 resistant E14 ES cell clones, 5 (out of 451) appear to carry a correctly targeted allele of *Mospd3*, corresponding to a targeting frequency of 1.1%.

3.2.2.10 Expansion and verification of targeted clones

Correctly targeted clones, of both CGR8 and E14 ES cells, were thawed from the frozen master plate and expanded under continuing G418 selection. The integrity of the gene targeting integration was examined in these clones using, both, the 3' probe to label the *SpeI* digests, and the 5' probe to detect *KpnI* restriction fragments, in order to analyse the upstream integration site (Figure 3.9 A and B). Only two of the CGR8 targeted clones exhibited the anticipated Southern blot pattern. E14 targeted clones are currently in the process of being verified (by Madina Kara).

To establish the absence of any unwanted targeting vector integrations in the genome of these targeted clones, the ES cells were also probed using a neomycin specific probe - in conjunction with an *NdeI* diagnostic digest of the integration site. Only a single fragment at the expected size of 17kb was found in both the clones (data not shown).

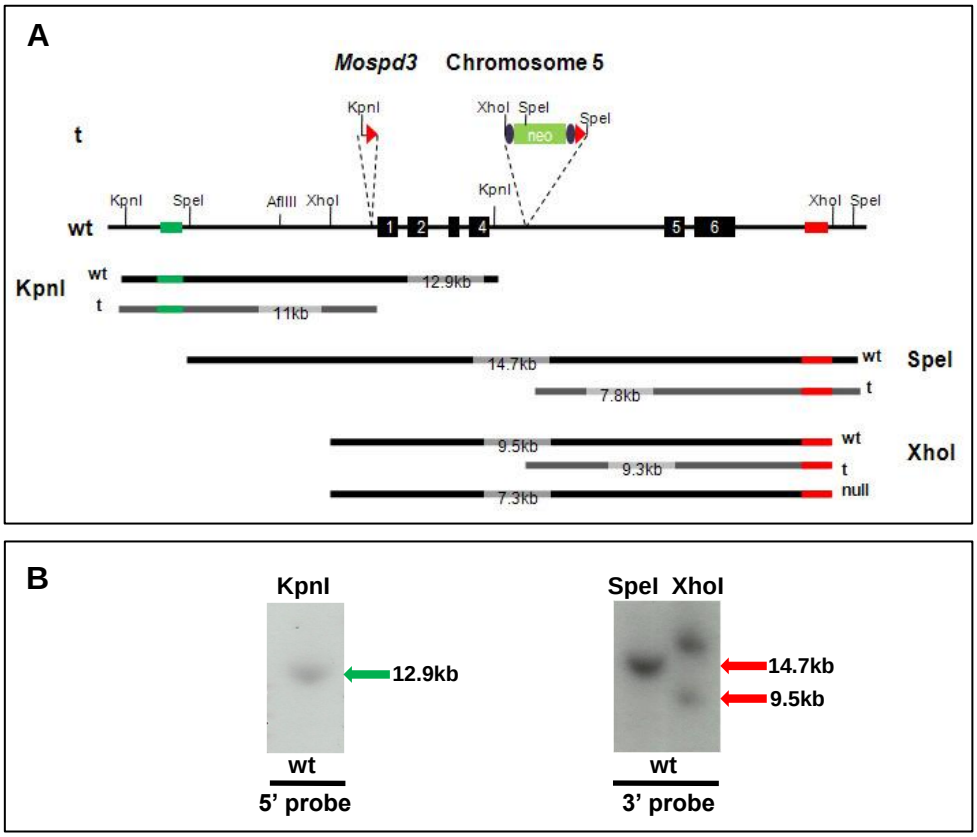


Figure 3.8: Design and testing of Southern blot screening strategy for *Mospd3* targeted clones. (A) Design of 5' and 3' probe hybridisation to different restriction enzyme fragments of genomic DNA in order to distinguish between wild type, targeted and null alleles in ES cells. The upper panel in (A) shows the wild type allele of *Mospd3* as well as the integration sites of the single *loxP* site (marked as red triangle) upstream of *Mospd3* exon1 and *FRT-neo-loxP* cassette (*FRT* site marked as purple oval) downstream of *Mospd3* exon 4 to generate the targeted allele. The lower panel of (A) shows the anticipated restriction fragments detected with 5' probe (green line) or 3' probe (red line) labelling of different restriction digests of genomic DNA. The solid black lines mark restriction fragments anticipated in wild type alleles (and the null/ *loxP* recombined allele in *XhoI* digests) whilst the grey lines mark the restriction fragments anticipated for correctly targeted alleles (t) of *Mospd3*. (B) Southern blot verification of anticipated wild type restriction fragment sizes using 5' and 3' probe labelling of wild type CGR8 genomic DNA. Abbreviations: neo, neomycin resistance marker; wt, wild type allele or wild type ES cell genomic DNA; t, targeted allele.

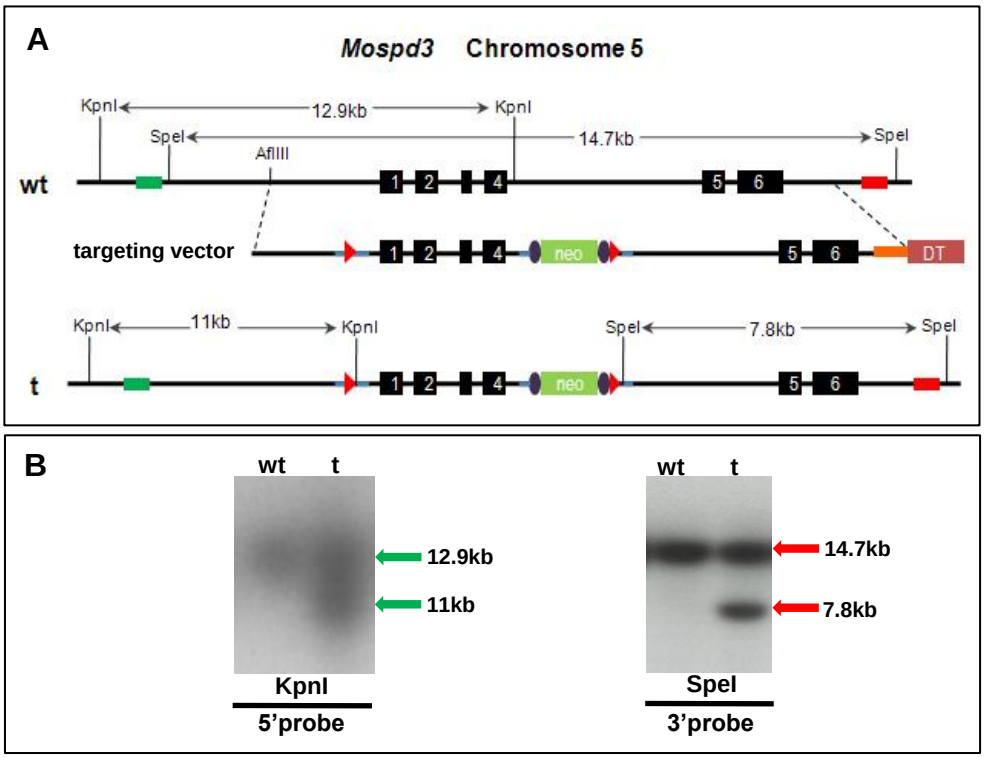


Figure 3.9: Southern blot analysis of *Mospd3* targeted ES cell clones.
(A) Schematic showing a map of the *Mospd3* genomic locus (wild type allele, wt), the targeting vector and the *Mospd3* targeted allele (t). The exons of *Mospd3* are represented by black boxes. The Southern blot probes used are marked by green boxes for the 5' probe and red boxes for the 3' probe. The targeting vector construct includes the relative position of the *loxP* site (red triangles), the *FRT–neo-FRT-loxP* cassette (*FRT*, purple ovals) and the diphtheria toxin (DT) cassette. (B) Southern blot analysis of ES cell genomic DNA to screen for ES cells carrying a correctly targeted allele of *Mospd3*. Abbreviations: DT, diphtheria toxin cassette; t, targeted allele or ES cell clone containing a correctly targeted allele of *Mospd3*, wt, wild type allele or ES cell clone carrying two wild type alleles of *Mospd3*.

3.3 DISCUSSION

We have demonstrated the successful creation of a f*lox*ed allele of, both, *Mospd1* and *Mospd3* in ES cells, as well as the subsequent generation of a *Mospd1* null allele. The construction of the conditional targeting vectors, for both genes, has been almost entirely performed by using a recombineering approach (Lee et al., 2001; Liu et al., 2003). Recombineering facilitates the fast and efficient construction of even complex targeting constructs. It has been estimated that the use of this approach can, on average, reduce the construction time of targeting constructs from several months, when using traditional restriction and ligation subcloning methods, to just a few weeks, when using a recombineering approach (Lee et al., 2001).

In agreement with these estimates, we found that the generation of the conditional targeting construct of *Mospd1* was quick and effortless, using a recombineering approach. We only used conventional restriction and ligation methods to subclone the *Mospd1* targeting construct, with sufficiently long homology arms, from a BAC into a high-copy plasmid backbone. All subsequent steps, leading to the generation of the f*lox*ed *Mospd1* targeting construct, were performed using recombineering methods. Using recombineering, it was possible to integrate both *loxP* sites into the targeting vector in exactly the positions chosen in the initial conditional allele design (Figure 3.1).

In contrast to *Mospd1,* the generation of the conditional targeting vector of *Mospd3* proved to be more time consuming. As indicated in section 3.2.2.3, it was not possible to subclone a suitably sized genomic fragment of *Mospd3*, including homology arms, by *Xba1* restriction and ligation. We struggled to find alternative restriction sites which would allow a conventional subcloning approach for the *Mospd3* targeting sequence.

We, therefore, chose to subclone the *Mospd3* targeting fragment from the BAC by gap repair (Lee et al., 2001; Liu et al., 2003). For this subcloning approach, homology sequences flanking the desired genomic fragment of *Mospd3* were PCR amplified and subcloned into the pNEB-DT target plasmid. Electroporation of recombination efficient EL350 cells (containing the *Mospd3*-BAC) with the linearized plasmid should lead to the retrieval of the *Mospd3* targeting sequence by homologous recombination.

This method, in particular the design of the homology regions, is highly dependent on knowing the correct genomic sequence of *Mospd3* in the BAC. Our initial attempts, at subcloning *Mospd3* by gap repair, failed. It later emerged that parts of the genomic sequence downstream of exon 6 of *Mospd3* (the last exon of *Mospd3*), which we had originally used to generate the initial 3' homology for gap repair, had been mis-annotated in the mouse genome build (NCBI m33).

Both publicly available mouse genome resources (NCBI and Celera) are created by a tiling approach of sequenced contigs, to stretch across whole *Mus musculus* genome. The genome sequence published by NCBI is updated yearly.

Analysis of the mouse genome build, NCBIm34, published in May 2005 revealed, that the initially designed 3' gap repair homology had been re-annotated to a location approximately 45kb away from the expected site: in the intergenic sequences immediately downstream of *Mospd3* exon 6. A subsequently designed homology arm resulted in the successful subcloning of the *Mospd3* target sequence.

The only way this problem might have been foreseen would have been by sequencing the entire sequence intended for subcloning. This would have been a time consuming and complicated endeavour as the intergenic sequences flanking *Mospd3* are quite large, and may be hard to sequence due to the abundance of repetitive elements in non-coding sequences. The design of homology sequences to target the *loxP* sites has the potential to be affected by similar problems. Interestingly, a recent high-throughput recombineering study, which generated more than 100 BAC targeting vectors, reported that in about 10% of the cases the generation of the vector failed due to BAC clones being wrongly mapped in the mouse genome, or chimaerisms and rearrangements occurring in the BAC clones (Poser et al., 2008).

Even though this sequence problem held up the construction of the *Mospd3* targeting vector, traditional vector construction methods would most likely have also been affected by the same sequence mis-annotation. In fact, one of the *XbaI* restriction sites initially chosen for *Mospd3* subcloning had been located in the same mis-annotated genomic sequence. Just like the 3' homology sequence designed for the gap repair, this *XbaI* site has also been re-annotated.

Table 3.1: Targeting frequencies of *Mospd1* and *Mospd3* conditional targeting vector in two different ES cell lines.

Cell line	Mospd1	Mospd3
CGR8	1.7%	1.5%
E14	2.2%	1.1%

Another hurdle, for the use of conventional cloning methods for constructing a conditional *Mospd3* targeting vector, was the absence of suitable restriction sites to allow such an approach. Analysis of the restriction enzyme pattern in the genomic sequence of *Mospd3* revealed a lack of suitable unique sites to allow the introduction of *loxP* sites into regions flanking exons 1 to 4 of *Mospd3*.

Once construction of the conditional targeting vectors of *Mospd1* and *Mospd3* had been concluded, they were electroporated into ES cells. The frequencies of homologous vector integrations in G418 resistant clones, from two different ES cell lines, ranged from 1.7% to 2.2% for the *Mospd1* targeting vector, and from 1.1% to 1.5% for the *Mospd3* targeting vector (Table 3.1).

The process of homologous recombination into ES cells is not very efficient, with targeting frequencies often less than 0.1 - 1% (Joyner, 1999). The targeting frequencies obtained for *Mospd1* and *Mospd3* are better than these average values. Several factors are known to enhance the rate of homologous recombination of a targeting vector in ES cells. These factors include the use of longer homology arms in the targeting vector as well as ensuring isogenicity between the targeting vector and the ES cell DNA (Deng and Capecchi, 1992; Hasty et al., 1991b; te Riele et al., 1992).

There are several ways in which the targeting frequencies of the constructs generated in this study could have been improved. One possible way would have been to increase the length of the homology arms. This could be achieved by using the whole BAC as the targeting construct (Valenzuela et al., 2003; Yang and Seed, 2003). Vector targeting frequencies of up to 8.3% (and 3.8% average) have been reported when using a BAC targeting vector (Valenzuela et al., 2003). The targeting frequency of such large vector constructs does not appear to be adversely affected by using non-isogenic vector DNA. The drawback of such an approach, however, is that screening for integrations at the correct locus cannot be performed by conventional Southern blotting methods.

Another way to improve the targeting frequency of the conditional knockout vectors of *Mospd1* and *Mopd3* would have been to use isogenic vector sequences. At the time, when the *Mospd1* and *Mospd3* targeting vectors were generated, the only BAC clones publicly available were generated from C57BL/6 mouse DNA, whilst

the ES cells used in this study were obtained from a 129/Ola mouse strain (Hooper et al., 1987; Nichols et al., 1990). As both, C57BL/6 and 129/Ola, mouse strains are only distantly related, there is likely to be differences in the genomic sequence between the *Mospd1* and *Mospd3* genes in the BAC targeting vectors, and the *Mospd1* and *Mospd3* alleles in the ES cells.

The reason for this mouse strain discrepancy was that, at the time, BAC clones from a 129 mouse strain were not publicly available. The first resource of BAC clones from a 129 mouse strain (129Sv) was reported in 2005 (Adams et al., 2005). Although 129Sv is a different substrain of 129 from the 129/Ola strain (used to derive E14 and CGR8 ES cells), the targeting frequency obtained using vectors generated from these BAC clones should be better. A Japanese research group has since reported the generation of BAC clones from 129/Ola cells (Ohtsuka et al., 2006). These clones are now available through the Japanese Collection of Research Bioresources (JCRB) Gene Bank.

Another approach, to ensure isogenicity between the targeting construct and the endogenous gene sequence, would have been to use C57BL/6-derived ES cells for gene targeting. C57BL/6-derived ES cells have not been widely used for gene targeting experiments. Compared to 129-derived ES cell strains, these ES cells were previously found to be more difficult to propagate *in vitro* as well as being less efficient at generating chimaeras and contributing to the germ line (Auerbach et al., 2000; Brook and Gardner, 1997). A report in 2003, comparing the efficiency of a C57BL/6 ES cell line versus 129 ES cell lines at generating targeted mutation in mice, found that their C57BL/6-derived ES cells performed only marginally less well in the generation of germ line transmitting chimaeras (Ware et al., 2003). The use of C57BL/6-derived ES cells, for the generation of transgenic mouse lines, has the added advantage of circumventing the need for laborious breeding to establish the targeted mutation on a homogeneous ($\geq$ 99%) inbred C57BL/6 background.

Another aspect in the generation of the *Mospd1* and *Mospd3* targeted allele which would benefit from improvement is the screening process for correct integrations of the conditional targeting vector. Southern blot screening of neomycin resistant clones can be time consuming, especially if the targeting frequency of the vector only ranges from 1.1% to 2.2%. The approach would benefit from a high-

throughput method of determining ES clones which carry an integration of the targeting vector at the desired locus. Two methods applicable for such a screen would be long range PCR, or the more recently developed loss-of-native-allele assay (Valenzuela et al., 2003). Although these methods enable a quick screen for targeted alleles, detailed analysis of the integration site by Southern blot is still necessary. During the integration of the conditional targeting vector of *Mospd1*, we found that in some of the neomycin resistant clones only part of the conditional targeting vector had integrated into the endogenous *Mospd1* locus. Southern blot analysis of these clones indicated that the *loxP* site in the intron between exons 1 and 2 of the *Mospd1* targeting vector had failed to integrate into the ES cell DNA. There is a chance that, due to its small size of only 100bp, the lack of this *loxP* site might not have been picked up by long range PCR. Similarly, unless the PCR primers used for the loss-of-native-allele assay were designed to the single *loxP* target site, and not the more likely target site of the *FRT-neo-FRT-loxP* cassette, the lack of the *loxP* sequences would not have been detected with this method, either.

CHAPTER 4: FUNCTIONAL STUDY OF TARGETED ES CELLS

4.1 INTRODUCTION

ES cells are pluripotent cells capable of populating all germ layers of a host blastocyst *in vivo*, including the germline. This ability to generate germline transmitting ES cell/blastocyst chimaeras facilitates the introduction of genetic modifications into the mouse model.

In this chapter, we describe our attempts to generate transgenic mice from ES cells carrying the conditional targeted alleles of *Mospd3* and *Mospd1* (which have been described in their creation in Chapter 3). In both conditional alleles, the upstream exons of the two *Mospd* genes are flanked by *loxP* sites so that Cre-mediated recombination results in the loss of the entire coding sequence of these genes. Tissue-specific and induced Cre-expression in mice carrying the conditional allele would allow us to study the effect of loss of function of both *Mospd* genes in a particular tissue or at a defined time point of development. In addition, the question of a possible genetic redundancy between the two *Mospd* genes could be answered by generating double knockout animals.

A complimentary approach, to studying the role of genes in cell lineage development in the knockout mouse model, is to employ *in vitro* differentiation of ES cells carrying the targeted mutation. Through culture in 3 dimensional aggregate structures, called embryoid bodies, ES cells can be induced to differentiate into cells of all three germ layers of the embryo. This process recapitulates early embryonic development. Specialized cell lineages, such as cardiomyocytes, can be rapidly derived through this *in vitro* system, facilitating initial studies of loss-of-function mutations on the differentiation and function of specific cell lineages, prior to generating a knockout mouse model. Both, *Mospd3* and *Mospd1* are expressed in the adult mouse heart. In order to determine whether *Mospd1* might be involved in heart development, and thus share functional redundancy with *Mospd3*, we chose to use the hemizygous *Mospd1* knockout ES cell lines, generated in Chapter 3, to study the effect of loss of *Mospd1* expression on the potential of ES cells to differentiate into cells of the cardiac lineage *in vitro*.

Unless specifically mentioned, all ES cell clones (carrying either a cko allele of *Mospd1* or *Mospd3*, or a *Mospd1* null allele) described in this chapter have been derived from CGR8 ES cells.

4.2 RESULTS

4.2.1 Generation of chimaeras

All five conditionally targeted clones of *Mospd1*, generated from a CGR8 parental ES cell line, were analysed for their growth in culture before using them for blastocyst injections. The two clones most closely resembling the growth pattern of wild type CGR8 ES cells were used to generate chimaeras. Blastocyst injections were kindly performed by Gail M^{c}Walter at the Genetic Intervention and Screening Technologies Service, as well as Jan Ure at the Institute of Stem Cell Research (both University of Edinburgh). Unfortunately, neither of these 2 conditionally targeted clones resulted in the generation of germline competent chimaeras. Clone 1 produced only 1 female and 1 male chimaera from 98 injected blastocysts and 30 pups born, whilst the second clone resulted in 4 male and 2 female chimaeras out of 33 injected blastocysts and 14 pups born. However, none of these 5 chimaeras showed a significant level of chimaerism - as observed by the agouti coat colour striping on the black (C57BL/6) host blastocyst-derived animals. The male chimaera generated from conditionally targeted *Mospd1* clone 1 was allowed to breed with two C57BL/6 females, but produced no transgenic offspring.

In the case of *Mospd3*, the two CGR8 ES cell clones carrying the correctly targeted allele of *Mospd3* were used for chimaera generation. From a total of 162 injected blastocysts and 19 pups born none yielded any chimaeric offspring.

4.2.2 Karyotyping

Due to the problems observed with the generation of chimaeras from blastocyst injection, karyotype analysis was performed on all ES cell clones carrying a conditional allele of either *Mospd1* or *Mospd3*. ES cell clones carrying the hemizygous null allele of *Mospd1* were also analysed for their karyotype stability. The karyotypes of the targeted clones were compared to their parental wild type ES cell lines (CGR8 and E14). Euploid murine somatic cells contain 40 chromosomes (2n=40) (Figure 4.1 A (left panel)).

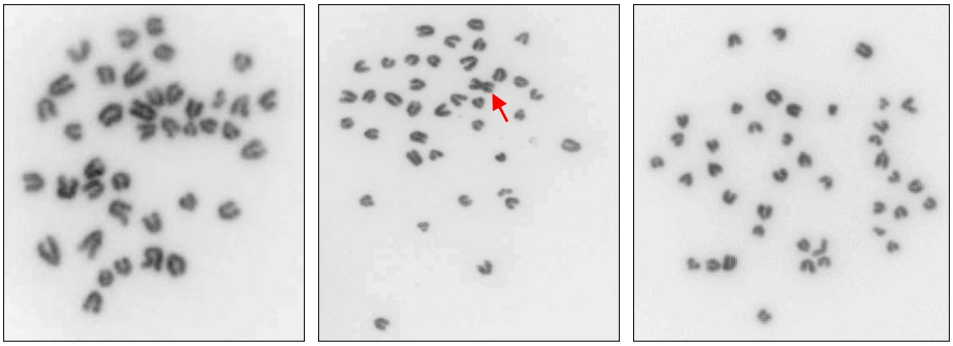
A count of the number of chromosomes present in 20 to 30 chromosome spreads per cell line revealed that CGR8 wild type ES cells and almost all the

targeted lines generated from this line had an abnormal karyotype (Figure 4.1 A and B). In 30% of CGR8 cells at passage 23, the presence of an extra chromosome (2n=41) was observed (see Figure 4.1 A (right-most panel), B). Among four of the *Mospd1* targeted ES cell clones generated from vector electroporation into these CGR8 cell at passage 17-19, we found that two exhibited a similar aneuploid pattern in the majority of the karyotyping spreads (*Mospd1*$^{0/cko}$ clones (3) and (4) in Figure 4.1 B). Both clones appear to have been generated from an integration of the targeting vector into CGR8 ES cells containing 41 chromosomes, as almost all the spreads counted had accumulated an extra chromosome. The same karyotype defect was also observed in the two CGR8 ES cell lines carrying the conditional allele of *Mospd3* (Figure 4.1 B, *Mospd3*$^{+/cko}$ clones (1) and (2)). Similar to *Mospd1*, it is likely that the *Mospd3* allele in these clones was targeted in CGR8 cells already carrying a trisomy, which was then clonally expanded. As G-banding analysis has not been performed, it is not known which chromosome has been duplicated, or if indeed the extra chromosome in the aneuploid targeted clones is the same as the one found in some of the CGR8 chromosome spreads.

In a small percentage (7%) of CGR8 cells at passage 21, one of the 40 murine acrocentric chromosomes has acquired two additional chromosome arms, making it appear metacentric (Figure 4.1 A (middle panel), B *Mospd1*$^{0/cko}$ clone (1)). The same type of chromosomal defect was also found in one of the *Mospd1* targeted clones, generated from the CGR8 parental cell line. Of all targeted CGR8 cell lines, only one (carrying a *Mospd1* conditional allele - Figure 4.1 A (left panel), B *Mospd1*$^{0/cko}$ clone (2)) appeared to have a normal karyotype.

The *Mospd1* null lines generated by Cre recombinase exposure of ES cell clones carrying a conditional allele of *Mospd1* (previously described in section 3.2.1.11) were also analysed for their karyotype. The *Mospd1* null ES cell lines exhibited the same karyotype pattern as the targeted cell line used to create them (Figure 4.1 B). Whilst the *Mospd1* null clones generated from the euploid *Mospd1*$^{0/cko}$ (clone 2) cell line also showed a normal karyotype, the second group of *Mospd1* null clones (generated from *Mospd1*$^{0/cko}$ clone 1) contained the same metacentric chromosome as its parental targeted cell line.

A



40 chromosomes | 40 chr/ 1 metacentric (X) | 41 chromosomes

B

wild type cell line	conditional targeted lines	null clones	40	X	41	other
CGR8			60	30	7	3
	***Mospd1*$^{0/cko}$** clone (1)		24	64	4	8
		***Mospd1*$^{0/-}$** two clones	26	61	4	9
	(2)		84	0	11	5
		***Mospd1*$^{0/-}$** two clones	80	0	5	15
	(3)		35	0	55	10
	(4)		43	0	48	9
	***Mospd3*$^{+/cko}$** clone (1)		39	0	53	8
	(2)		48	0	42	10

Figure 4.1: Karyotype abnormalities in targeted ES cell clones derived from CGR8 cells. (A) This panel shows three predominant karyotypes observed in CGR8 and clones derived from this ES cell line. From left to right these panels show an euploid karyotype spread containing 40 chromosomes, a euploid karyotype featuring an abnormal metacentric chromosome (highlighted by the red arrow) and an aneuploid karyotype of 41 chromosomes (trisomy). The table in B) shows the karyotyping results obtained from each of the targeted *Mospd1* and *Mospd3* ES cell clones, generated from CGR8 wild type ES cells. The clone identity numbers are shown in brackets. Karyotyping results are also shown for *Mospd1* null (*Mospd1*$^{0/-}$) clones generated by Cre-mediated recombination in *Mospd1*$^{0/cko}$ clones (1) and (2). The results obtained from the *Mospd1* null clones correspond to the average distribution of karyotypes observed in two *Mospd1*$^{0/-}$ clones. The values listed correspond to percentages of cell with a particular karyotype in all the karyotype spreads of a particular ES cell clone. The karyotype categories in the table correspond to the karyotypes panels pictured in (A). (X) denotes spreads containing a metacentric chromosome. (Other) denotes karyotypes not within these categories including chromosome counts of less than 40 or more than 41 chromosomes. Abbreviations: Chr, chromosomes

In contrast to the chromosome abnormalities observed in CGR8 cells, the other ES cell line (E14), routinely used in this laboratory, was found to contain a normal number of chromosomes in 97% of all the chromosome spreads analysed. All targeted ES cell lines, carrying a conditional allele of either *Mospd1* or *Mospd3,* derived from E14 cells were found to be euploid (data not shown). The karyotype analysis of targeted E14 cell lines was performed with the help of Madina Kara.

4.2.3 Self renewal and differentiation potential of *Mospd1* null ES cells

Prior to studying the effect of loss of expression of *Mospd1* on cardiomyocyte differentiation, we confirmed that the loss of function of this gene did not interfere with the potential of ES cells to self renew or initiate differentiation. For this purpose, a standard self renewal and differentiation assays was performed by plating 1,000 cells (either wild type ES or *Mospd1* null) into each well of a 6-well culture dish, in ES cell medium. After 24 hours, the culture medium in half the wells was replaced with ES cell medium lacking LIF (Leukaemia Inhibitory Factor). LIF is used as a growth supplement in murine ES cell culture to keep cells in an undifferentiated, self-renewing state. Removal of LIF drives the cells towards differentiation. Cells were maintained in these conditions for 3 days to form colonies, which were then stained for alkaline phosphatase activity. Undifferentiated ES cells exhibit high levels of alkaline phosphatase activity, which diminishes as cells start to differentiate.

The stained colonies were subsequently scored for their level of differentiation. Tight red staining colonies were scored as undifferentiated stem cell colonies, whilst colonies with a fainter more pink stain or colonies with a red centre and unstained differentiated periphery were counted as mixed colonies. If there was no red staining detected in a colony, it was scored as differentiated.

Under self renewing conditions in the presence of LIF, the contribution of stem cell, mixed and differentiated colonies appeared to be similar between the two wild type ES cell lines (CGR8 and E14) and the two karyotypically normal *Mospd1* null ES cells clones used in this assay, as indicated by chi-squared analysis (Figure 4.2 B). This result suggests that the loss of *Mospd1* in the null cells does not have an

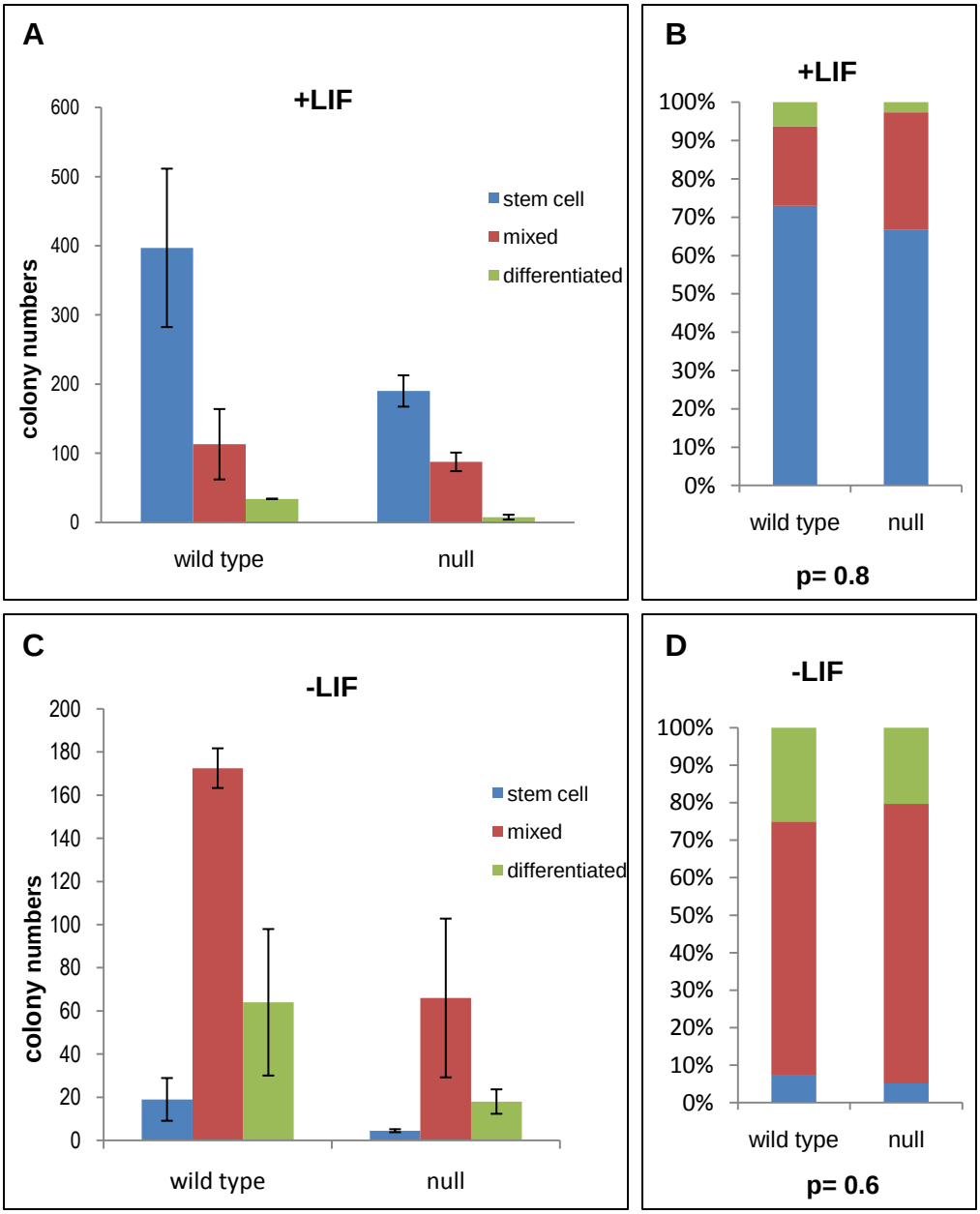


Figure 4.2: Self-renewal and early differentiation of *Mospd1* null ES cells compared to wild type ES cell lines (E14 and CGR8). (A) Graph shows the number of colonies of either undifferentiated (stem cell), mixed and differentiated phenotype formed by wild type or *Mospd1* null ES cells in culturing conditions promoting stem cell self renewal (+LIF) whilst graph (B) represents the distribution of the three different colony type in the total number of colonies as a percentage. (C) This graph shows the numbers of stem cell, mixed and differentiated colonies in both wild type and Mospd1 null ES cells in conditions, which promote ES cell differentiation (-LIF), whilst graph (D) shows the contribution of the three different colony types in the total number of colonies formed under these differentiating conditions. Chi-squared analysis of the colony type contribution (B and C) in +LIF and –LIF conditions indicates that the contribution of stem cell, mixed and differentiated colonies does not differ significantly between *Mospd1* null ES cells and wild type cells. Error bars in (A) and (B) correspond to standard deviation.

effect on the ability of these ES cells to self-renew in the same way as wild type cells. However, we did detect lower overall colony numbers in the *Mospd1* null clones compared to the two wild type ES cell lines (Figure 4.2 A). A similar picture was seen under differentiating conditions, where the ratio of the three colony types appeared to be distributed comparable to wild type clones (Figure 4.2 D), but overall fewer colonies were seen in the wells (Figure 4.2 C). As under self renewing conditions, the comparable contribution of stem cell, mixed and differentiated colonies, between wild type and *Mospd1* null cells, in the absence of LIF indicates that the loss of *Mospd1* does not have an effect on the potential of ES cells to commence differentiation.

It is difficult to account for the lower total colony number observed in the *Mospd1* null colonies compared to the two wild type ES cell clones. As only three different *Mospd1* null clones have been used for this comparison, and all these clones have originated from the same conditionally targeted CGR8-derived ES cell clone, it is likely that this is a clonal effect. Due to the limited availability of karyotypically normal *Mospd1* null clones, the sample number in this experiment was very low. This cardiomyocyte differentiation experiment would, no doubt, benefit from being repeated once *Mospd1* null clones have been generated from karyotypically stabile E14 ES cells, to confirm the results gained in this study.

As the loss of *Mospd1* did not appear to affect the early differentiation potential of ES cells, the *Mospd1* null clones were analysed for their potential to differentiate into functional cardiomyocytes. In this experiment, *Mospd1* null ES cells were allowed to form embryoid body aggregates in hanging drops. These 3-dimensional structures have been found to mimic *in vivo* embryo development, enabling the ES cells in the EB to initiate differentiation *in vitro* into all three germ layers. After harvesting, the embryoid bodies (EBs) were cultured for a further 5 days in ES cell medium lacking LIF before each EB is added to one well of a pre-gelatinised 24-well plate. Under these conditions, the embryoid bodies attach to the gelatine surface and flatten. Over the course of 9 days in culture, the formation of beating cardiomyocytes was observed. The cardiomyocyte differentiation was represented in this experiment as a percentage of the number of EBs exhibiting beating cardiomyocyte activity, over the number of embryoid bodies present in a 24-

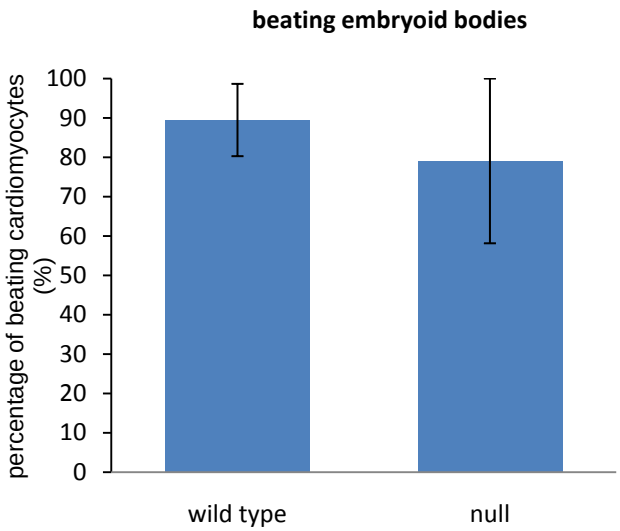


Figure 4.3: Comparison of the *in vitro* cardiomyocyte differentiation potential of wild type and *Mospd1* null ES cells. The graph shows the percentage embryoid bodies forming beating cardiomyocytes with respect to the number of embryoid bodies (EBs) plated out on gelatine. Data represents results from two wild type ES cell lines (E14 and CGR8) and three *Mospd1* null clones. All plated EBs were monitored throughout the nine day differentiation course. Error bars = standard deviation.

well culture dish (Figure 4.3). This was done because some wells contained two EBs, or not enough EBs were obtained from the 5 day EB suspension culture to fill all 24 wells of the plate.

Although this experiment has only been performed once, using two wild type ES cell lines (CGR8 and E14) and three *Mospd1* null clones, the result strongly indicates that the lack of *Mospd1* does not have any effect on the ability of the cells to differentiate into cardiomyocytes. 89% of the embryoid bodies derived from wild type ES cells, and 79% from *Mospd1* null clones formed beating cardiomyocytes. Thus, the cardiac-specific differentiation potentials of these cells appear to be comparable.

We also tried to discern whether the loss of *Mospd1* may affect the function of the cardiomyocytes, rather than their formation. As a measure of their function, we chose to analyse the contraction frequency of myocardial foci in the embryoid bodies on day nine of cardiomyocyte differentiation. The number of contractions per minute, of 5 randomly chosen beating EBs per wild type and *Mospd1* null ES cell clones, was counted under the microscope. We found that the rate of contraction varied greatly even amongst beating EBs of the same cell line, leading us to conclude that contraction frequencies would not be a suitable measure to determine functional differences between *Mospd1* null and wild type ES cell lines.

4.3 DISCUSSION

As *Mospd1* is expressed in the mouse heart, and is suspected to be a possible genetic modifier affecting the phenotype in the *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals, a *Mospd1* null ES cell line was generated to assess the effect of lack of this gene on the *in vitro* cardiac differentiation potential of these cells. We found that ES cells lacking Mospd1 expression were still able to self renew and differentiate into cells of the cardiac lineage. Because the cardiac defect in *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals was only observed in d19 embryos, it is still possible that *Mospd1* may only compensate for loss of *Mospd3* at later stages of cardiac development, rather than the early cardiac differentiation mimicked by embryoid bodies in *in vitro* culture. In order to determine whether the function of these *in vitro* derived cardiomyocytes might be affected by the loss of *Mospd1*, we looked for changes in the rate of contraction in the beating EBs.

In a study on the effect of loss of β_1-integrin on cardiomyocyte formation, it was found that the absence of this extracellular matrix interactor protein led to an increase in the beating frequency of cardiomyocytes in EBs (Fassler et al., 1996). However, the rate of contraction of beating cardiomyocyte is very heterogeneous, in correspondence with the heterogeneity of different myocardial cell types present in these beating EB cultures (Boheler et al., 2002). The frequency of contractions in wild type cardiomyocyte cultures also changes with progressive differentiation of cardiomyocytes, from slow contractions of early pacemaker cells to the faster contractions of intermediate differentiated myocardial cells. Upon terminal cardiomyocyte differentiation, contractions become slower or even stop.

Additional experiments to study the cardiomyocyte differentiation potential of *Mospd1* null ES cells could include testing the beating embryoid body cultures for cardiac muscle specific proteins, such as cardiac myosin heavy chain (MHC), to determine the percentage of cells which have developed into cardiac muscle.

It is also possible to isolated single cardiomyocyte cells from beating embryoid body cultures. These ES cell derived cardiomyocytes can be analysed for their electrophysiological properties, as well as the expression of cardiac gene products by RT-PCR and immunohistochemical analysis. To determine if Mospd proteins are involved with cell adhesion or tethering of the contractile apparatus, the

localisation of components of the sarcomeric cytoskeleton as well as desmosomal and adherens junction proteins could be examined in isolated cardiomyocytes.

In this study, ES cells containing a targeted allele of *Mospd1* and *Mospd3* were used to generate chimaeric animals by blastocyst injection, in order to produce a transgenic conditional mouse line of either gene. The generation of chimaeras failed in the case of *Mospd3*, and even if chimaeras were produced, as for two of the *Mospd1* clones, the level of ES cell contribution to the animals was very low. In addition, one male chimaera bred with two C57BL/6 females did not produce any transgenic offspring.

An explanation for these low numbers of chimaeras produced, was gained from subsequent karyotyping analysis of the targeted ES cell clones generated from the CGR8 parental ES cell line. Almost all the clones had acquired a trisomy of one of the chromosomes, which was also found in a proportion of the wild type CGR8 cells.

It has been shown, in a number of reports, that mouse ES cells are prone to accumulation of chromosome abnormalities in long term *in vitro* culture (Liu et al., 1997; Longo et al., 1997; Robertson, 1987). Of 540 mouse ES cell lines analysed in a Japanese study, only 60% exhibited a normal karyotype (Sugawara et al., 2006). Karyotype abnormalities in mouse ES cells include monosomy, trisomy, polyploidy, as well as translocation and duplication of chromosomal regions (Rebuzzini et al., 2008). The most commonly observed karyotype abnormality, in mouse ES cells, is trisomy of chromosome 8 (Liu et al., 1997). Trisomy of chromosome 11 is also quite common (Sugawara et al., 2006). It is believed that the presence of an extra chromosome 8 or 11 gives these cells a growth advantage, which leads to a rapid replacement of a euploid ES cell, in culture, with aneuploid cells (Liu et al., 1997; Sugawara et al., 2006). This thinking corresponds with findings that euploidy in an ES cell culture can deteriorate from 100% to 16% in only 5 passages (Nichols et al., 1990).

The CGR8 cells used to generate the *Mospd1* and *Mospd3* conditional allele were electroporated with the targeting vector at passage 17 to 19, and then clonally expanded. This expansion holds the risk of selecting a particular cell, which may have already acquired a karyotype defect. In fact, CGR8 cells at passage number 19

were found to already contain as many as 30% aneuploid cells. Karyotype instability is a common problem with all of the established and widely distributed mouse ES cell lines; as, by their nature, they have been in culture for a long time. Some cell lines have been shown to be more stable over long periods of culture, even past passage number 33, whilst others can accumulate aneuploid cells at very early passages (Longo et al., 1997; Nichols et al., 1990).

Several studies have found that an abnormal karyotype is the main cause for failure of ES cells to generate germline-transmitting chimaeras (Liu et al., 1997; Longo et al., 1997; Nagy et al., 1993). The higher the number of aneuploid cells in an ES cell culture, the fewer blastocyst injections of these cells resulted in chimaeras being generated. Also, the level of ES cell contribution to the chimaera was reduced. If the contribution of euploid cells in a given ES cell culture dropped below 50%, no chimaeras could be generated at all.

Interestingly, the reason for using CGR8 ES cells for the generation of the conditional knockout allele of *Mospd3* and *Mospd1* was the finding that this cell line performed particularly well when used for blastocyst injection, resulting in a high number of germline transmitting chimaeras (Nichols et al., 1990) (personal communication Jan Ure, Edinburgh). However, over the years of continuous culture the CGR8 line appears to have lost this potential, along with an increase of karyotype instability in these cells.

Only one of the conditional targeted clones generated in the CGR8 parental cell line resulted in the production of a suitable number of chimaeras. Similar to one of the ES cell clones studied by (Longo et al., 1997), this clone shows no obvious karyotype defects with nearly all of the cells being euploid (*Mospd1*$^{0/cko}$ (2), Figure 4.1B). But despite their normal chromosome contribution, the cells appear to have lost their totipotency as they show only a very low level of ES cell contribution, and the male ES cells were not able to lead to a sex distortion in the chimaeras towards male animals. Both, sex distortion and the level of ES cell contribution have been shown to influence the ability of a clone to contribute to the germline. These findings indicate that this clone may have lost the potential to be transmitted in the germline, either due to a subtle genetic defect not detected by karyotyping, or a due to partial

commitment of the cell, precluding them from contribution to all germ layers of the embryo.

Karyotype analysis of E14 cells at passage 22 and 23, alongside the CGR8 cells and targeted clones, revealed that about 97% of the E14 ES cells were euploid. In addition, a recent gene trap mutagenesis project in our laboratory, using the E14 wild type line, found that these cells were capable of generating a good level of germline transmitting chimaeras. It was therefore decided to use E14 cells in the attempt to retarget the *Mospd1* and *Mospd3* allele (Chapter 3). To date, 5 clones containing the *Mospd3* conditional allele, and 4 clones targeted at the *Mospd1* locus have been identified. All 9 clones display a more than 95% euploid karyotype, and are anticipated to result in the successful generation of a transgenic mouse line of both, *Mospd1* and *Mospd3*.

CHAPTER 5: GENERATION AND TESTING OF ANTI- MOSPD1 AND ANTI-MOSPD3 ANTIBODIES

5.1 INTRODUCTION

Mospd3 is part of a novel gene family in vertebrates. To date, the only evidence for a function of *Mospd* genes has come from the analysis of a gene trap integration into the mouse *Mospd3* gene. Loss of function of this mammalian specific *Mospd* gene resulted in a cardiac defect, and in severe cases death, in neonatal mice. The severity of this phenotype appeared to be affected by a genetic modifier, possibly another member of the *Mospd* gene family such as *Mospd1*. Little else is known about the function of this gene family, although the *Mospd3* gene trap phenotype and the protein domain similarities of Mospd3 to an ancestral protein (Major Sperm Protein) in *C. elegans*, indicate that Mospd3 protein may be involved in the development or structural integrity of the right ventricular chamber of the heart.

To test this hypothesis as well as learn more about this family of genes, it is important to study the function of the proteins encoded by members of the *Mospd* gene family. An important tool for the functional study of novel proteins is the use of antibodies. Antibodies could be used to study the tissue-specific distribution as well as the subcellular localisation of both Mospd3 and Mospd1. This would help to confirm whether Mospd proteins form part of the desmosome or adherence junction complexes, which provide structural integrity in the heart. In addition, specific αMospd antibodies can be used to identify novel protein interactors of Mospd3 or Mospd1.

Various strategies were available to generate the antigens required for raising specific antibodies against Mospd3 and Mospd1. One option would have been to clone the appropriate cDNA into an expression vector in-frame with a tag, such as His_6 or GST. Overexpression of the fusion construct in protease deficient *E.coli* cells would allow the production of a tagged Mospd protein, which could be purified by immuno-chromatography using sephadex or charged nickel columns. However, due to the high levels of protein conservation between the MSP and transmembrane domains of Mospd3 and Mospd1, some of the antibodies raised against the whole proteins would most likely display antibody cross-reactivity.

To overcome this problem, we used an alternative approach for generating immunogenic antigens. Instead of using the whole protein to raise antibodies, we

chose to design a small peptide antigen. Small peptide antigens have become the preferred immunogen class used today, as they can be quickly and easily synthesised. Publicly available computational sequence analysis programs[7] can be used for the identification of suitable peptide epitopes.

Peptide immunogens, which typically consist of 12 to 20 amino acid residues, can be designed to reduce unwanted cross-reactivity with related proteins. When designing a suitable peptide immunogen for Mospd1 and Mospd3, we chose to avoid peptide sequences within the MSP and transmembrane domains of either protein as these regions exhibit the highest degree of sequence conservation (see Chapter 1 figure 1.2). The most N-terminal protein sequences of Mospd1 and Mospd3, on the other hand, show no significant sequence similarities making potential cross-reactivity between antibodies, raised against either protein, unlikely. An added advantage of peptides designed to either the N- or C-terminal protein sequences is that these regions often exhibit low structural complexity and are less likely to be buried in the tertiary structure of the native protein. Peptide antibodies may not recognise epitope sequences buried within the tertiary structure of a protein. Also, antibodies raised against surface exposed peptide sequences are likely to detect both the native and the denatured form of a protein.

In this study, peptide antigens were designed to the most N-terminal 20 amino acid sequences of Mospd1 and Mospd3 (Figure 5.1). The antibodies raised against these peptide sequences were anticipated to bind specifically to either Mospd1 or Mospd3, and to not exhibit any cross-reactivity between these two proteins.

[7] http://bio.dfci.harvard.edu/Tools/antigenic.pl

5.2 RESULTS

5.2.1 Generation of peptide antigens and raising of polyclonal αMospd1 and αMospd3 antibodies

Specific peptides were designed and synthesised by Thistle Research. The sequence of these peptides corresponded to the most N-terminal 20 amino acids of either Mospd1 or Mospd3, as these sequences showed minimal sequence identity between the two proteins (Figure 5.1).

Both peptides, conjugated to KLH carrier proteins, were subsequently used for the inoculation of two rabbits per peptide (performed by PTU/BS (Roslin)). Preimmune sera, as well as first, second and third bleeds (taken at weeks 4, 8 and 12, respectively) were collected.

5.2.2 Testing of antiserum by ELISA and western blotting

ELISA (Enzyme-Linked ImmunoSorbent Assay) was performed using preimmune and week 4 to 12 sera, in order to determine whether antibodies detecting Mospd1 or Mospd3 were present in the antisera, and to find a good antiserum sample for further purification and immunoblot analysis.

The ELISA involved the immobilisation of the peptide antigen onto a solid support and subsequent detection of this peptide by antibodies present in the antisera. As pre-immune serum was taken from the rabbits prior to inoculation with the peptide antigen, it should not contain those antibodies which would specifically detect the immobilised peptide. The absorbance reading (as a measure of antibody binding) of pre-immune serum could therefore serve a baseline of non-specific antibody binding (Figure 5.2). An increase in the absorbance reading in subsequent bleeds indicated a specific immune response to the peptide antigen in the rabbits. With further booster injections, the amount of specific antibody present in the antisera should further increase between week 4 and 12, resulting in an increase of absorbance readings. This increase in the levels of specific antibody binding was observed in the antisera samples of one rabbit inoculated with the Mospd3 peptide (Figure 5.2 A) and one rabbit inoculated with the Mospd1 peptide (Figure 5.2 D).

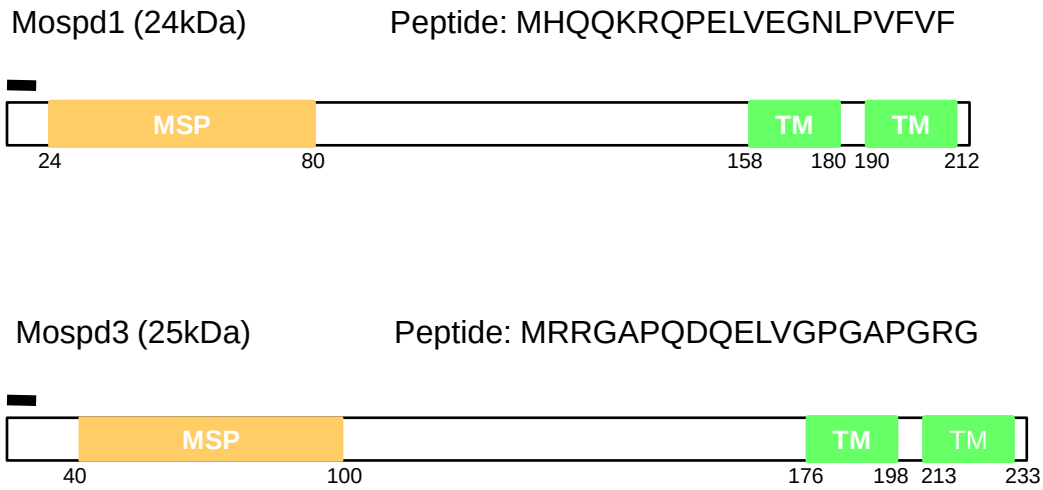


Figure 5.1: Location of protein sequences used to generate synthetic peptide antigens. The sequence (▬) used for generation of the synthetic peptide and subsequent raising of Mospd1 and Mospd3 specific antibodies is located at the N-terminus of both Mospd proteins. The N-terminal protein sequences preceding the MSP domain correspond to non-conserved sequences between the Mospd1 and Mospd3 protein (also see Figure 1.2. for protein sequence alignment of both Mospd proteins).

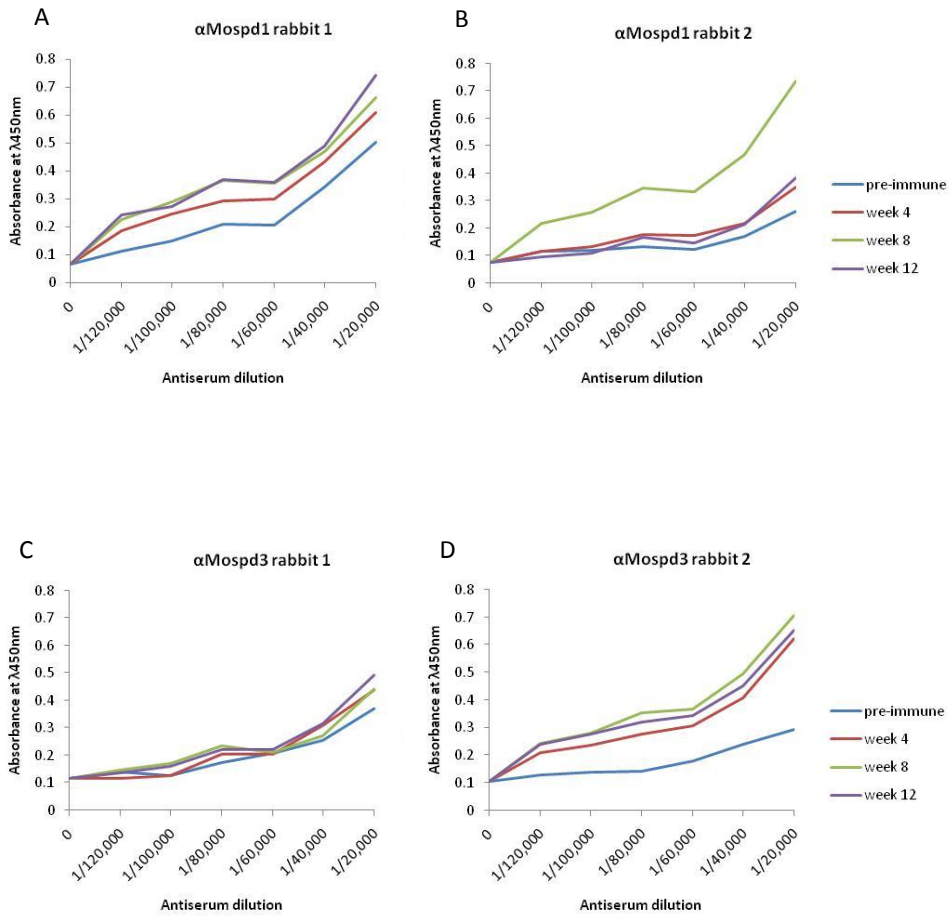


Figure 5.2: ELISA of αMospd1 and αMospd3 rabbit antisera. αMospd1 rabbit antisera from two rabbits (A, B) as well as αMospd3 antisera from two rabbits (C, D) were analysed for antibody binding to the relevant peptide antigen using ELISA . The assay measured the amount of specific antibody binding of preimmune serum and antisera obtained at week 4, 8 and 12 at various antisera dilutions. The level of antibody binding to the immobilised peptide antigen is expressed as a measure of the absorbance at λ450nm. An increase in the slope of the curve at week 4, 8 or 12 compared to the preimmune curve indicates the presence of specific antibodies against Mospd1 or Mospd3.

The ELISA results from rabbit 1 injected with Mospd3 peptide show barely any increase in the antibody binding activity between pre-immune and any of the later bleeds (Figure 5.2 C). This could indicate either an insufficient immune response in this rabbit or, possibly, the production of specific antibodies with weak affinity to Mospd3.

In the second rabbit immunised with Mospd1 peptide, on the other hand, the concentration of specific antibody rises in the sera from week 4 and 8, but drops drastically in the last bleed (Figure 5.2 B). Peptide antigen booster injection are performed, not only to increase the amount of specific antibody produced, but also to promote immunoglobulin shift from IgM to IgG antibodies, which usually display stronger epitope binding affinity. In some cases, however, the resulting antibodies can show reduced binding affinity or specificity (Roitt and Delves, 2001). This might have been the case in the second rabbit immunised with the Mospd1 peptide. Another explanation could be an infection in this rabbit after the second bleed. Based on the ELISA results, we chose to use week 8 antiserum for further immunoblotting analysis of the antibodies, as this bleed, from all four rabbits, appeared to give consistently good results.

In order to analyse the specificity of the antibodies, western blot analysis was performed on COS7 cells transiently transfected with either a *Mospd1*-cDNA, or a cDNA encoding an eGFP-Mospd3 fusion protein, in a mammalian expression vector. The *Mospd1* expression construct (Appendix figure 6) was obtained as an I.M.A.G.E clone from the I.M.A.G.E Consortium[8], whilst the *eGFP-Mospd3* fusion vector was produced by *EcoRI* digestion and ligation of the *Mospd3*-cDNA into a pC2-EGFP vector by Dr. Richard Axton, Edinburgh (both pEGFP vector maps in Appendix figure 5). Western blot membranes of protein lysates from these transfected COS7 cells were probed with αMospd1 and αMospd3 preimmune and week 8 sera. Whilst the pre-immune serum did not show any signal on the western blot (Figure 5.3), the week 8 αMospd1 antiserum (used at a dilution of 1/1,000) detected a protein band at approximately 24kDa corresponding to the predicted molecular weight of the Mospd1 protein.

[8] http://image.llnl.gov

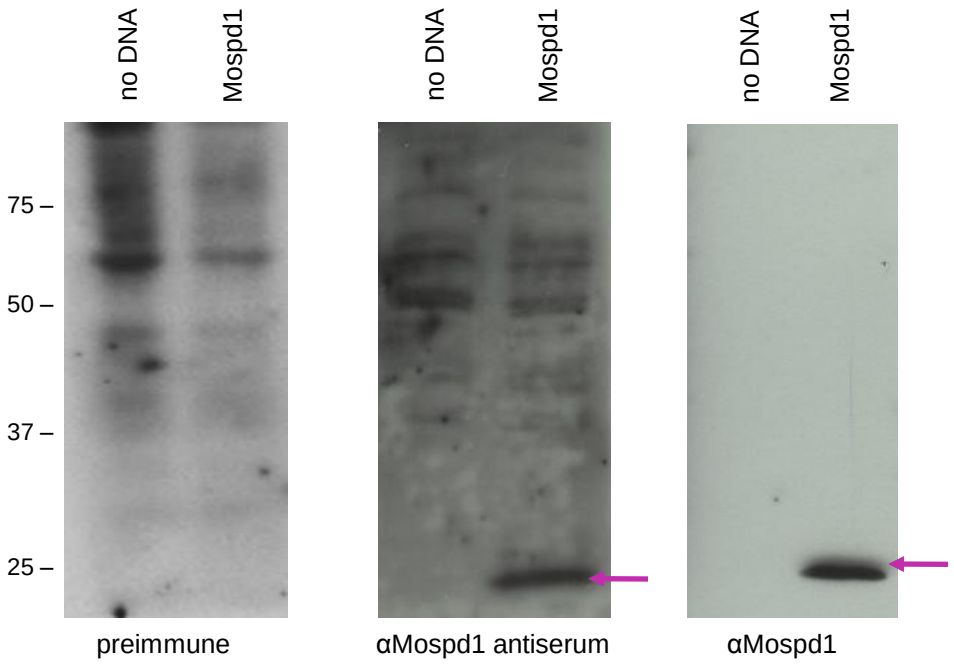


Figure 5.3: Assessment of polyclonal αMospd1 antibody purification from rabbit antiserum. Preimmune serum, αMospd1 antiserum and immunoaffinity purified polyclonal αMospd1 antibody were used for western blot analysis of cell lysates from COS7 cells transfected with either no DNA or a Mospd1 expression construct . The purple arrow marks the 24kDa Mospd1 protein expressed by transfected COS7 cells.

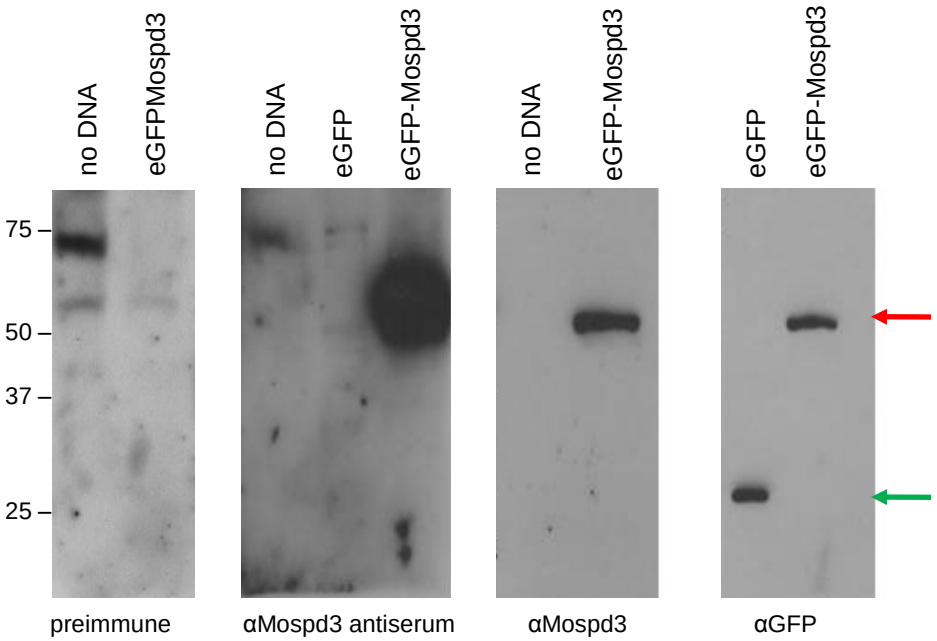


Figure 5.4: Assessment of polyclonal αMospd3 antibody purification from rabbit antiserum. Preimmune serum, αMospd3 antiserum, immuno-purified polyclonal αMospd3 antibody and αGFP were used for western blot analysis of cell lysates from COS7 cells transfected with either no DNA, an eGFP expression vector or an eGFP-Mospd3 fusion construct. The red arrow indicates the 52kDa eGFP fusion protein and the green arrow marks the 27kDa eGFP protein expressed by transfected COS7 cells.

The αMospd3 antiserum also detected a specific band (Figure 5.4) in the lysate of transfected COS7 cells. This band, at 52kDa, corresponded to the eGFP-Mospd3 fusion protein. The identity of the fusion protein was confirmed using a commercially available monoclonal anti-eGFP antibody (Figure 5.4).

Although using crude week 8 antisera for western blot analysis enabled the detection of both Mospd1 and Mospd3 in COS7 cell lysates, a high level of non-specific background signal was detected, despite increasing the timing of washing steps significantly compared to standard western blotting protocols. Neither reduction of the amount of COS7 protein sample or antisera concentration (1/2,500, minimum concentration able to generate signal) nor the use of different blocking agents resulted in clearing the non-specific background signal (data not shown).

5.2.3 Purification of αMospd1 and αMospd3 antibodies from week 8 antiserum

Immunoaffinity purification of the polyclonal αMospd antibodies from week 8 antiserum was performed in an attempt to increase specificity.

Initial purification attempts involving the use of a Protein 'A' column, which binds immunoglobulins via interactions with their heavy chain, were unsuccessful (data not shown).

A second approach, to purify the antibodies from the serum, involved the use of immunoaffinity columns containing the immobilised Mospd1 or Mospd3 peptide antigen which was used to inoculate the rabbits for antiserum production. These columns were supplied by Thistle Research. The αMospd1 and αMospd3 antibodies purified by this method were tested on western blots of protein lysates from transiently transfected COS7 cells, and from mouse tissues.

5.2.4 Testing of immunoaffinity purified polyclonal αMospd1 and αMospd3 antibodies on transfected COS7 cell lysates

Probing western blot membranes of transfected COS7 cell lysates with the immunoaffinity purified antibodies showed that the non-specific background has been significantly reduced. Used at a dilution of 1 in 200, the purified αMospd3

antibody detected the 52kDa protein band corresponding to the eGFP-Mospd3 fusion protein (Figure 5.4). The identity of this fusion protein was again confirmed using an anti-eGFP antibody.

At the same 1 in 200 dilution, the immunoaffinity purified αMospd1 antibody detected a distinct protein band at 24kDa, the predicted molecular weight of Mospd1 protein (Figure 5.3). The identity of this protein has not been confirmed as we have not been able, to date, to clone an eGFP-Mospd1 fusion construct.

We assessed whether the antibodies we had raised were specific to either one of the Mospd proteins. We detected no protein bands when αMospd1 antibodies were incubated with western blots of *eGFP-Mospd3* transfected COS7 cells, nor when αMospd3 antibodies were incubated with blots of Mospd1 transfected COS7 cells (Figure 5.5).

These experiments confirmed that we have successfully raised polyclonal antibodies, which were able to detect Mospd1 and Mospd3. We have also shown that, using specific peptide antigens corresponding to non-conserved regions of Mospd1 and Mospd3, we were able to raise antibodies which can distinguish between these two closely related Mospd proteins.

5.2.5 Testing of polyclonal αMospd1 and αMospd3 antibodies on mouse tissue lysates

In order to establish whether the purified antibodies were able to detect endogenous mouse Mospd1 and Mospd3, they were used to probe western blots of mouse tissue lysates of various organs, including heart and kidney (αMospd1, Figure 5.6).

Even at very short exposure times a multitude of protein bands were detected throughout the lanes of the western blot. The amount of non-specific signal was too high to determine specific protein bands at 24kDa or 25kDa, corresponding to endogenous mouse Mospd1 and Mospd3, respectively (Figures 5.6, data for αMospd3 not shown). Reducing the amount of protein sample or antibody concentration did not help to reduce the non-specific background.

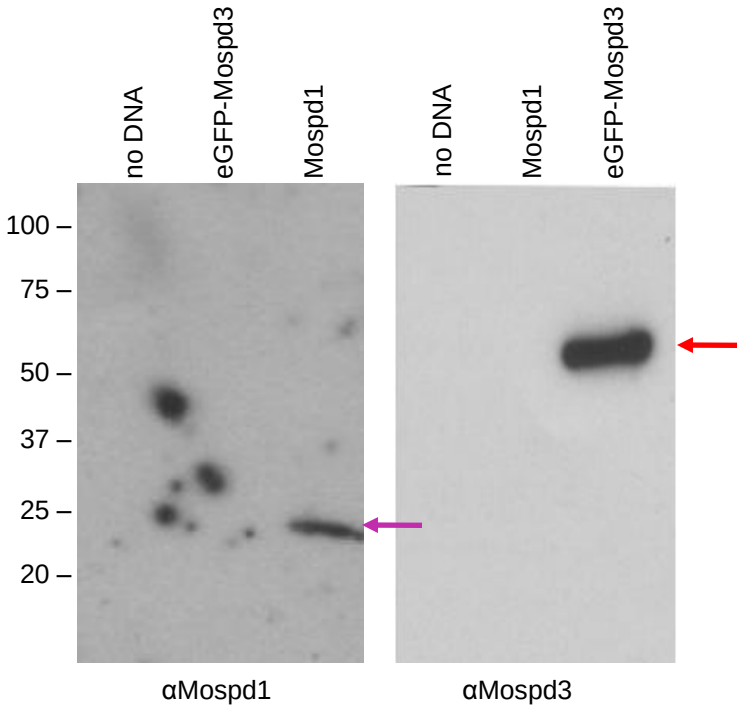


Figure 5.5: Western blot analysis to assess cross-reactivity between immunoaffinity purified polyclonal αMospd1 and αMospd3 antibody. Antibody cross-reactivity was assess by antibody labelling of protein lysates of COS7 cells transfected with either an eGFP-Mospd3 fusion construct or a Mospd1 expression construct. The red arrow indicates the eGFP-fusion protein and the purple arrow marks the Mospd1 protein expressed in transfected COS7 cells.

In order to identify the Mospd3 specific protein band in mouse tissue lysates, heart and kidney protein extracts of wild type animals were compared to those from homozygous *Mospd3*$^{Gt1lmf/Gt1lmf}$ gene trap animals which do not express Mospd3. However, the non-specific background signal made it impossible to distinguish a particular protein band present in wild type lysates but absent in *Mospd3*$^{Gt1lmf/Gt1lmf}$ samples (data not shown).

5.2.6 Reduction of background signals from mouse tissue western blots

In order to reduce the background signals observed on mouse tissue western blots, the purified antibodies were incubated with mouse serum prior to being used on wild type mouse tissue lysates. This pre-incubation might have reduced the background signals if they had been due to non-specific antibody binding to circulating murine proteins. However, in this case, no reduction of the background signals was observed.

An alternative approach to eliminate the background signals involved pre-incubation of the purified antibodies with ES cell lysates which did not express the Mospd protein of interest. Using TritonX100 protein lysates of *Mospd1* null ES cells (described in section 3.2.1.11), the purified αMospd1 antibody was pre-incubated and subsequently used to probe western blot of mouse heart and kidney protein lysates. This approach resulted in a reduction of the non-specific background signal. The same result was achieved by pre-incubation of the αMospd1 antibody with wild type CGR8 ES cell lysates.

In fact, after pre-incubation with either cell lysate, a protein signal more distinct than most of the surrounding background bands was detected in heart and kidney sample. This protein band, with an approximate molecular weight of 24kDa, appeared to correspond with endogenous murine Mospd1. If this signal indeed corresponds to endogenous Mospd1, it would also indicate that this protein is not expressed in undifferentiated ES cells (as both *Mospd1* null and CGR8 wild type lysates appeared to reduce western blot background signals but not the 24kDa band).

Further confirmation of the identity of this signal would be required. This could be done by out-competing the αMospd1 antibody with its peptide antigen prior to probing western blots of mouse tissue lysates.

It was not possible to perform similar pre-incubation experiments with the purified αMospd3 antibody as, at the time, neither *Mospd3* null ES cell lines nor *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals were available for tritonX100 extraction of proteins.

These preclearing (pre-incubation with ES cell lysates) experiments show that it was possible to reduce the non-specific binding of the αMospd1 antibody. However, the presence of persistent non-specific background indicated that, even after preclearing with *Mospd1* null ES cell lysates or *Mospd3*$^{Gt1lmf/Gt1lmf}$ mouse tissue lysates, the antibodies were unlikely to be suitable for use in immunohistochemical procedures. Due to the remaining non-specific binding activity of the αMospd1 antibody, it would not be possible to correlate a positive immunohistochemical staining pattern with specific detection of Mospd1 alone.

5.2.7 Generation of monoclonal αMospd1 and αMospd3 antibodies

As the polyclonal antibodies appeared to be unsuitable for functional analysis of Mospd proteins, monoclonal antibodies specific to Mospd1 and Mospd3 were generated. These monoclonal antibodies were raised against the same N-terminal 20 amino acid peptide antigen of Mospd1 and Mospd3 as the polyclonal antibodies. The peptides, synthesised and conjugated to KLH by Yorkshire Biosciences, were injected into mice. Splenocytes and lymph node cells of these inoculated mice were subsequently fused to myeloma cells to generate hybridomas expressing monoclonal αMospd1 or αMospd3 antibodies. The conditioned medium from positive hybridoma lines was then collected (Nelson et al., 2000; Shepherd and Dean, 2000).

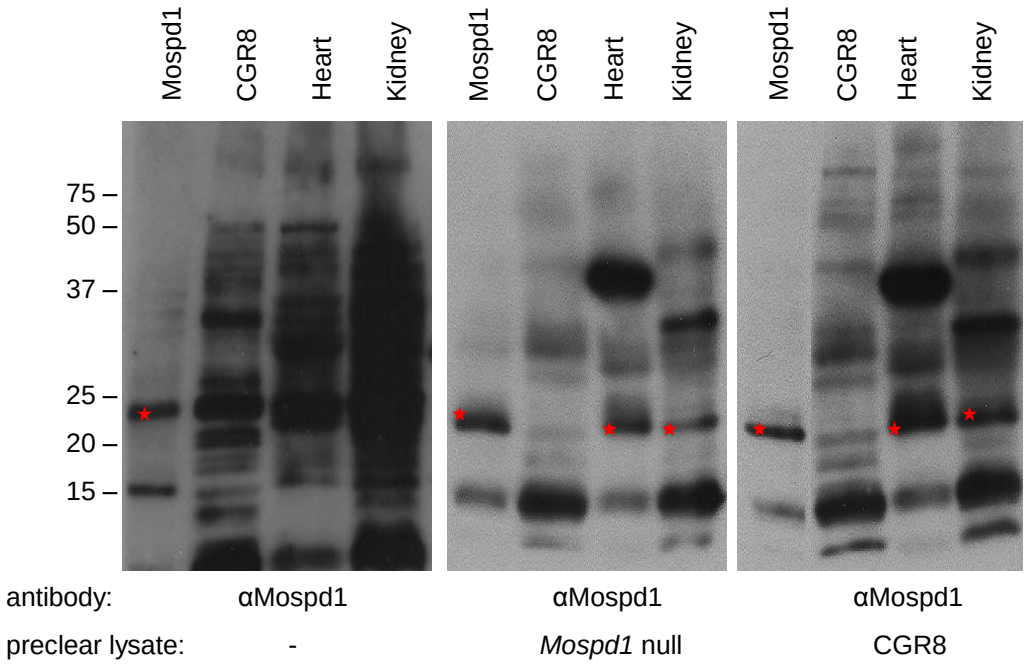


Figure 5.6: Western blot analysis of CGR8 ES cells and heart and kidney tissue lysates from mouse using polyclonal αMospd1 antibody. The immunoaffinity purified polyclonal was precleared by prior incubation with TritonX100 lysates of either Mospd1 null ES cells or CGR8 cells. Red asterisk indicate the proposed endogenous Mospd1 protein at a predicted molecular weight of 24kDa.

5.2.8 Testing of monoclonal antibodies on recombinant and endogenous Mospd proteins

Initial analysis and testing has to date only been performed on αMospd3 monoclonal antibodies with the help of Madina Kara. Conditioned medium containing αMospd3 monoclonal antibodies was used to probe western blots of COS7 cells expressing the eGFP-Mospd3 fusion protein (Figure 5.7). The monoclonal antibody detected a protein band of approximately 52kDa (eGFP-Mospd3), indicating that this antibody is able to detect the Mospd3 protein.

In subsequent western blot analysis of mouse tissue lysates of heart and kidney, the antibody detected a specific protein band at about 25kDa which could correspond with endogenous murine Mospd3 (Figure 5.7). In addition to this signal, two further protein bands were labelled (at approximately 15kDa and 50kDa).

Further comparison of the protein bands labelled in wild type and *Mospd3*$^{Gt1lmf/Gt1lmf}$ mouse tissue lysates, as well as outcompeting αMospd3 monoclonal antibody with the peptide antigen prior to western blot hybridisation, should help confirm the identity of these proteins.

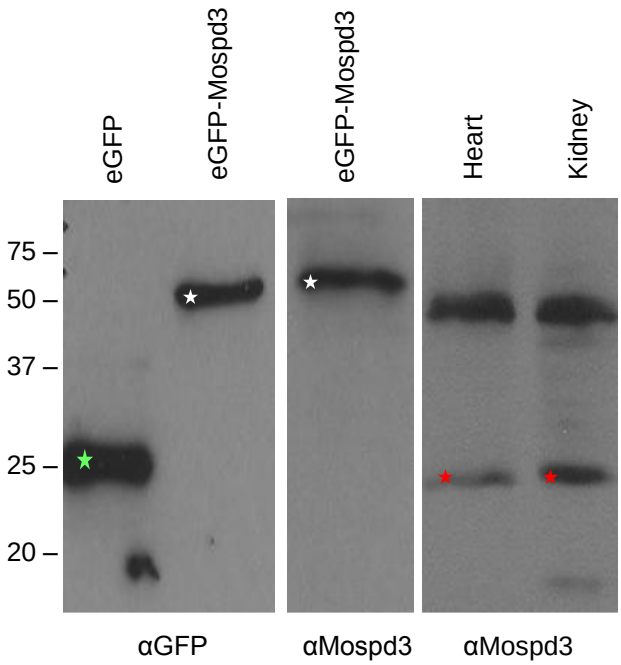


Figure 5.7: A monoclonal αMospd3 antibody detects protein in lysates from mouse heart and kidney tissues, and transiently transfected COS7 cells. Monoclonal αMospd3 and αGFP antibody were used to label lysates of COS7 cells transfected with eGFP alone or eGFP-Mospd3 fusion construct. Red asterisks indicate the proposed endogenous mouse Mospd3. The white and green asterisks mark the eGFP-Mospd3 fusion protein and the eGFP protein, respectively.

5.3 DISCUSSION

Antibodies represent a very powerful tool for the functional analysis of a gene. Not only can they provide detailed information about the cell and tissue specific localisation of the protein encoded by a gene of interest, but co-immunoprecipitation using antibodies can also facilitate the discovery of interacting proteins.

This thesis describes the generation, testing and purification of both polyclonal and monoclonal antibodies raised against N-terminal peptide antigens of Mospd1 and Mospd3. These antibodies were found to detect both recombinant and endogenous Mospd1 and Mospd3, as well as being able to successfully distinguish between these two closely related proteins.

Immunoblotting analysis of purified and precleared αMospd3 polyclonal antibodies, and αMospd1 monoclonal antibodies, detected expression of both Mospd3 and Mospd1 in adult mouse heart and kidney tissues. This is the first time the expression of Mospd proteins has been shown. The presence of both Mospd1 and Mospd3 in the same adult mouse tissues appears to be evidence for the hypothesis that both proteins perform a similar function and may be able to compensate for each other. This finding could also explain the phenotype variability in *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals (Chapter 1, section 1.1.3), as well as the gradual loss of the neonatal death phenotype in homozygote gene trap animals noted in offspring after about generation 12. It is possible that, through successive breeding, animals with improved survival and compensation by Mospd1 have been selected.

In addition to potential Mospd1 expression in adult heart and kidney, immunoblotting using αMospd1 indicated that Mospd1 may not be expressed in undifferentiated ES cells (Figure 5.6). Therefore, any differences observed in the physiology of undifferentiated *Mospd1* null ES cells (described in Chapter 4) might not have resulted from the loss of Mospd1, but might be more likely to be caused by clonal differences or an experimental effect due to the low numbers of clones analysed.

We have shown by immunoblotting that Mospd1 is expressed in mature heart cells but (most likely) not in ES cells. This indicates that at some point during cardiac development the expression of Mospd1 is initiated. To study the importance

of the *Mospd1* gene during cardiac development, we have used an *in vitro* cardiomyocyte assay (Chapter 4, section 4.2.3) which mimics the *in vivo* model. In light of the protein expression pattern of Mospd1 shown in this chapter, there appear to be two different explanations for the lack of phenotype observed in *Mospd1* null cardiomyocyte cultures. Either Mospd1 is not expressed and required in the earlier stages of cardiomyocyte differentiation analysed in this assay, or alternatively Mospd3 can compensate for the loss of Mospd1 in these cells. From results in this chapter, both appear to be valid conclusions. The answer to this question will require further immunohistochemical analysis of Mospd proteins during myocardial development both *in vitro* and *in vivo*.

Even though the polyclonal αMospd1 and αMospd3 antibodies were shown to be able to detect recombinant Mospd proteins in COS7 cell lysates and potentially endogenous Mospd in mouse tissue lysates, these antibodies could not be sufficiently purified. This may be due to the intrinsic nature of the polyclonal antibodies which contain multiple different antigen recognition sites. Up to 90% to 99% of antiserum could potentially contain antibodies which are not specific to the antigen of interest and can cause non-specific background cross-reactivity (Harlow and Lane, 1988). Even after purification using a specific peptide antigen affinity column, the polyclonal antibodies may still bind to other murine proteins, through antigen recognition sites not specific to Mospd1 or Mospd3. Nevertheless, the background of non-specific proteins detected, in our experiments, was unexpectedly high. These background problems are not uncommon due to the complexity of the immune reaction generated in animals inoculated to raise the antibodies.

Unlike polyclonal antibodies, monoclonal antibodies represent a more homogenous population of immunoglobulins which are derived from one clonal hybridoma line and thus contain the same antigen recognition site (Nelson et al., 2000). This feature makes these antibodies more specific and, therefore, preferred primary antibody reagents. Our initial western blotting analysis indicated that the monoclonal αMospd3 antibody strongly detects not only recombinant but also endogenous protein.

Pending further analysis to verify that the proteins detected by these monoclonal antibodies are indeed Mospd3 or Mospd1, and additional purification of the monoclonal antibodies from the hybridoma supernatant, these antibodies can be used to determine whether Mospd proteins form part of the cell adhesion complexes at the intercalated disc.

CHAPTER 6: SUMMARY AND PROSPECTIVES

6.1 SUMMARY

Mospd3 has previously been identified in a gene trap experiment. The resulting loss of function in *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals led to neonatal death of some animals, whilst the survivors exhibited cardiac abnormalities. This is the only published study relating to the function of the mammalian *Mospd* gene family. Genetic background variability in the observed phenotype suggested the existence of a genetic modifier locus. Whilst there are likely to be a whole host of possible genetic modifiers which could be the reason for the compensation of the *Mospd3* gene trap phenotype, we chose to analyse the closely related *Mospd1* gene. The proteins encoded by *Mospd1* and *Mospd3* are very similar in the identity and arrangement and of their functional domains, which show a high degree of conservation between both proteins.

The proteins encoded by the *Mospd* family of genes contain an N-terminal MSP domain as well as one or two C-terminal transmembrane domains. Whilst the presence of the transmembrane domain suggests a membrane-bound protein, the only indication concerning the function of the MSP domain comes from the nematode major sperm protein (MSP). Similarity with this protein, which appears to be an ancestral form of the MSP domain (Tarr and Scott, 2005) suggests that Mospd proteins may function as mediators of protein-protein interactions.

In fact, in *C. elegans*, MSP was found to be essential for actin-like cell motility of nematode sperm. Further analysis of this nematode system indicated that the sperm cells do not only highly express MSP protein but also LIM and PDZ domain proteins, which are known to associate with actin-filaments to mediate cross-linking and tethering of the actin cytoskeleton.

This information along with the cardiac phenotype of *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals, which display a thinning of the right ventricular myocardium and a possible defect in the structural integrity and function of the heart muscle, led us to speculate that Mospd proteins may be involved in the cell adhesion between the cardiomyocytes of the right ventricular heart wall. Such speculations have been substantiated by similar defects in animal models and humans, involving a loss of function of known cell-adhesion components such as proteins associated with

desmosomes and adherens junctions (Awad et al., 2008; Kostetskii et al., 2005; Perriard et al., 2003).

Alternatively, it is possible that Mospd proteins are involved in vesicle trafficking as they are similar in structure to VAP proteins. The VAP protein class also evolved from the nematode MSP protein and consists of an N-terminal MSP domain and a C-terminal transmembrane domain protein (Tarr and Scott, 2005). A preliminary study of the subcellular localisation of the eGFP-Mospd3 fusion construct in COS7 cells, showed a punctate fluorescence signal akin to vesicle-specific labelling, which would support this possibility (Forrester, unpublished data). Vesicle trafficking is a process which ensures correct distribution of molecules not only within the cell, but also between neighbouring cells. The central nervous system was found to be particularly sensitive to disturbances in this process, and almost all human syndromes caused by vesicle trafficking defects appear to display a neurological component (Olkkonen and Ikonen, 2006). One of these human syndromes, called Danon disease caused by a mutation in lysosomal-associated membrane protein 2 (Lamp2), is characterized clinically by hypertrophic cardiomyopathy, skeletal myopathy and variable degrees of mental retardation (Yang and Vatta, 2007). Danon disease is, as yet, the only described vesicle trafficking related defect which also display cardiac abnormalities. However, this disease affects multiple organs and we have not observed any obvious muscle wasting or behavioural defects in *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals.

An alternative explanation for the punctuate fluorescence of eGFP-Mospd3 in transfected COS7 kidney cells could be that the protein localises to focal adhesion sites of the cells with the matrix or gelatine culture surface. Interestingly, a recent review on the function of α-actinin reports that this evolutionary conserved actin-crosslinking protein localises to stress fibres and lamellopodia; as well as cell-cell and cell-matrix adhesion sites in non-muscle cells, such as kidney cells (Otey and Carpen, 2004). A punctuate localisation pattern has also been reported for the α-actinin-associated LIM domain protein (ALP) in spherical chick cardiomyocytes in primary culture (Henderson et al., 2003). ALP was found to co-localise with α-actinin at focal matrix adhesion sites prior to the formation of the sarcomeric cytoskeleton (myofibrillogenesis) in these myocardial cells. Once myofibrillogenesis

has occurred in these cardiac cells, both ALP and α-actinin localise to the sarcomeric Z-disc. The similarity in the localisation patterns of Alp, α-actinin and eGFP-Mospd3 in non-muscle and pre-sarcomeric cardiomyocytes indicates that Mospd3 may be able to interact with α-actinin, and maybe Alp.

The third possible reason for the punctate fluorescent signal could be that the transfected COS7 cells (derived from monkey kidneys) are sequestering the overexpressed foreign eGFP-Mospd3 fusion protein in lysosomal vesicles for degradation. An answer to whether Mospd3 is involved in cell-cell and cell-matrix adhesion, or vesicle trafficking will, no doubt, be reached through future subcellular immunohistochemical analysis of this protein.

To date, other than the conclusions drawn from the similarity of the gene trap phenotype of *Mospd3* to other loss of function mutations in adherens junction and desmosomal proteins, there is little functional information about this gene. In addition, the expression pattern of *Mospd3* is not very informative in determining gene function. Unlike the initial *β-gal* reporter gene expression of the trapped *Mospd3* locus, which seemed to indicate specific expression of *Mospd3* in the heart of developing mouse embryos, subsequent RT-PCR and *in situ* hybridisation analysis suggested that the gene might be expressed much more widely in the developing and adult mouse (McClive et al., 1998; Pall et al., 2004). *Mospd3* expression has been detected in *in vitro* differentiated EBs, hematopoietic tissues of the mouse embryo; as well as a range of tissues - including the heart - in the neonatal animal. This widespread expression pattern is corroborated by data from the Gene Expression Omnibus website.

A similar widespread pattern of expression has also been found for *Mospd1*. Like *Mospd3*, this gene is expressed in *in vitro* differentiated embryos at day 3 and 5; as well as in a range of embryonal hematopoietic tissues and in the adult mouse heart (data not shown). This overlapping expression pattern of *Mospd3* and *Mospd1*, particularly in the mouse heart, suggests that it could be possible for Mospd1 to compensate for the loss of Mospd3. In order to answer the question of a possible genetic redundancy between these two genes, and to analyse our hypothesis that Mospd3 and possibly also Mospd1 could be involved in providing structural integrity

in the heart, we have in this project generated the tools to facilitate a detailed functional study of both *Mospd* genes and their protein products.

Firstly, by employing a recombineering approach, conditional knockout constructs were generated for *Mospd3* and *Mospd1*. Both constructs, containing *loxP* sites which flank the MSP domain encoding exons of either *Mospd* gene, were successfully used to generate targeted alleles in ES cells. Recombineering proved to be an efficient technique for the generation of targeting vectors for both *Mospd1* and *Mospd3*. The generation of these targeting vectors by conventional restriction and ligation methods would most likely have proven more difficult. This would have been the case especially for *Mospd3*, as the genomic sequence of this gene was found to contain very few suitable restriction sites for this task. As recombineering does not require the use of restriction sites, it was possible to place the required *loxP* sites and targeting cassettes into the genomic sequences of our choosing. The only hurdle, we encountered using this approach, involved the mis-annotation of genomic sequences in the M*ospd3*-containing bacterial artificial chromosome (BAC). As mouse genome sequences are reviewed on a yearly basis, this issue is almost negligible for most current recombineering projects.

One of the most time consuming factors in the generation of the conditional knockout alleles of *Mospd1* and *Mospd3* was the targeting into ES cells and subsequent verification of correctly targeted clones. The use of whole BAC targeting vectors and novel innovative ES cell screening approaches (Valenzuela et al., 2003) has helped to make high-throughput targeting in ES cells feasible. The development of these high-throughput methods (recombineering and loss-of native-allele ES cell screening) led to the establishment of three major initiatives with the goal of large scale mutagenesis of all annotated genes in the mouse genome. The aim of the Knock-Out Mouse Project (KOMP), European Conditional Mouse Mutagenesis project (EUCOMM) and the North American Conditional Mouse Mutagenesis project (NorCOMM) is to saturate the mouse genome with targeted gene mutations. A recent search of the publicly available repository of targeted clones, held at the Sanger Centre[9], revealed the availability of an ES cell line carrying a conditional frame shift allele of *Mospd1*. The conditional allele generated by KOMP features an

[9] www.sanger.ac.uk/htgt/welcome

FRT-flanked promoterless targeting cassette containing a splice acceptor, *β-gal* reporter gene, neomycin resistance marker and polyadenylation site in the intron between exons 3 and 4 of *Mospd1*; as well as two *loxP* sites flanking exon 4 of the gene. The design of this allele allows the expression of gene sequences downstream of exon 3 of *Mospd1* to be disrupted due to a frame shift in the targeting cassette. As exon 4 contains a large part of the MSP domain encoded by *Mospd1*, the resulting protein will be, most likely, inactive. Subsequent use of flp recombinase allows the gene to be reactivated by deleting the promoterless cassette, whilst the use of Cre recombinase allows gene inactivation to be re-established. The use of this reactivatable allele of *Mospd1* may be useful, in combination with our own conditional allele, for future functional analysis of the *Mospd1* gene. There is no conditional allele available for *Mospd3* through KOMP, as the generation of a conditional targeting vector for this gene has been annotated as failed.

We generated E14 ES cells carrying targeted alleles of *Mospd1* and *Mospd3.* These cells were confirmed to be karyotypically normal and can now be used to generate chimaeric animals and establish transgenic mouse lines.

We have shown that the loss of *Mospd1* expression in ES cells had no effect on the ability of these cells to self-renew, nor did it interfere with their ability to differentiate into cells of the cardiomyocyte lineage. This finding suggests that *Mospd1* is not required for early cardiac development.

Specific polyclonal and monoclonal antibodies have been generated, which can detect, as well as distinguish between, Mospd1 and Mospd3 proteins. These antibodies have been used to show that, both, Mospd1 and Mospd3 are expressed in adult mouse heart and kidney. We have also shown that the Mospd1 protein is not expressed in undifferentiated ES cells, a finding that would account for the normal self-renewal and early differentiation potential in *Mospd1* null ES cells.

Unfortunately, the polyclonal antibodies could not be sufficiently purified to be used for further immunohistochemical analysis of Mospd1 and Mospd3, but initial results for the monoclonal antibodies are very promising. The use of these antibodies should facilitate detailed analysis of Mospd protein expression in the embryo and adult animal.

6.2 PROSPECTIVES

6.2.1 Protein analysis

Pending further validation and purification, the monoclonal antibodies can be used for detailed functional analysis of Mospd1 and Mospd3 proteins. These studies would include immunohistochemical studies, of whole-embryo and adult tissue sections, to determine where in the animal each Mospd protein is expressed. It would also be of interest to perform a detailed study of the localisation of Mospd3 in the heart. This would help to establish if the cardiac defect observed in particular in the right ventricle of *Mospd3*$^{Gt1lmf/Gt1lmf}$ animals stems from a loss of the protein specifically in the wall of this heart chamber.

To test our hypothesis that Mospd3, and possibly also Mospd1, might be involved in cell-cell adhesion at the intercalated disc of cardiomyocytes, subcellular localisation studies will be performed; and the distribution of either Mospd protein will be compared to the subcellular localisation of known desmosome and adherens junction (AJ) components such as desmoplakin, plakophilin-2, N-cadherin and ALP in cardiac muscle sections. Furthermore, multi-coloured immunohistochemical analysis will be performed to directly co-localise proteins that are predicted to interact. These co-localisation experiments could be conducted on heart muscle sections as well as *in vitro* differentiated cardiomyocyte cultures (derived from ES cells); along with cultured primary mouse cardiomyocyte and keratinocyte cells. Both isolated cardiomyocytes and cultured keratinocytes express high levels of desmosome and adherens junction proteins (Buxton and Magee, 1992; Westfall et al., 1997). Aside from their use in co-immunohistochemical analysis, these cells can also be used in co-immunoprecipitation experiments to determine whether Mospd proteins can interact with cell adhesion complex proteins, and to identify novel interacting proteins. If Mospd1 or Mospd3 do indeed interact with desmosomes or adherens junctions, it should be possible, using αMospd1 and αMospd3 antibodies, to co-immunoprecipitate components of such these adhesion complexes if bound by Mospd proteins, from cardiomyocyte or keratinocyte lysates. If the co-precipitated proteins correspond to known cell adhesion complex components, they can be identified by western blot analysis using antibodies against desmosomal or AJ

proteins, whilst mass spectrometry of the interacting proteins could be used to establish the identity of any novel co-precipitated proteins.

An alternative method to establish whether Mospd1 or Mospd3 form part of the desmosomal complex, in particular, entails western blot analysis of tritonX100 lysates of either keratinocytes or *in vitro* derived cardiomyocytes. It has been previously reported that desmosome complexes cannot be disassociated in tritonX100 cell lysis buffer (Cheng et al., 2004). Thus monoclonal antibody detection of Mospd1 or Mospd3 in the insoluble fraction of such lysates would indicate that either of these proteins are part of the desmosome complex.

Cardiomyocyte cells derived from *Mospd1* and *Mospd3* null ES cells as well as tissue sections and primary keratinocyte cell lines derived from *Mospd1* and *Mospd3* null transgenic animals (generated from the conditional knockout ES cell lines described in Chapter 3), will serve as valuable controls in all proposed protein function experiments. In addition, western blot analysis of these *Mospd* null animal tissues and cell lines could reveal whether there is any functional compensation for the loss of one Mospd protein by increased expression of the other.

If the monoclonal antibodies should at any stage prove to be unsuitable for any of the previously described protein localisation and immunoprecipitation studies, a recently developed alternative method to study the function of proteins without the need for specific antibodies can be employed. This novel method is based on the introduction of a fluorescent reporter tag into a gene of interest using a high-throughput BAC recombineering approach (Poser et al., 2008). The advantage of a BAC, compared to traditional gene targeting vectors or cDNA-based transgenes, is that due to its large size of 100 to 200kb the BAC targeting construct will almost always contain not only the entire sequence of the gene to be tagged but also endogenous regulatory sequences. This allows the BAC transgene, once integrated by homologous recombination, to be expressed at near-physiological levels.

The reporter gene tag used in this approach corresponds to a modified LAP tag (Cheeseman and Desai, 2005). This modular protein tag consists of enhanced GFP, an S-peptide sequence as well as a protease cleavage site, and can be used for tagging either the N- or C-terminus of a protein. The expression of this LAP tag in mammalian cells allows not only *in vivo* localisation analysis of a protein of interest;

but, using Protein ‘A’ immunochromatography, the protein can be isolated from cell lysates along with other proteins or DNA molecules it directly interacts with.

A particularly interesting aspect of this approach is that it can be used in ES cells. These ES cells containing the tagged construct of either *Mospd1* or *Mospd3* could in the first instance be differentiated into cardiomyocytes. This way, the subcellular localisation of Mospd proteins can be studied in these *in vitro* cardiac muscle cells and compared to the localisation pattern of desmosomal and AJ complex proteins. Moreover, the ES cells expressing tagged Mospd1 or Mospd3 could be used to generate transgenic mouse lines, thus allowing detailed protein functional analysis throughout the murine development and in the adult animal. Additionally, primary cell lines, which express the tagged Mospd protein of interest at high levels, could be isolated from mouse tissues and then used for live imaging of endogenous Mospd protein. These primary cell lines would also be ideal for co-localisation studies as well as co-precipitation of tagged Mospd1 and Mospd3, and their protein interactors.

6.2.2 *In vitro* studies in isolated cardiomyocytes

As previously mentioned (in section 6.1), in a transient transfection experiment in COS7 kidney cells, the eGFP-labelled Mospd3 protein displayed a punctuate signal (Forrester, unpublished). This pattern could potentially correspond to a co-localisation of Mospd3 with α-actinin protein at cell-matrix adhesion sites. In order to confirm whether Mospd3 can indeed interact with α-actinin, or as hypothesised with other desmosomal and adherens junction proteins, the *eGFP-Mospd3* plasmid could be transiently transfected into cardiomyocyte cells isolated from *in vitro* ES differentiation cultures or mouse hearts. If Mospd3 is involved in cell-cell adhesion, or structural integrity of the sarcomeric cytoskeleton with the cell membrane, we would expect the fluorescently labelled protein to localise to the intercalated discs of these myocardial cells. If (like Alp) Mospd3 can interact with α-actinin, a fluorescent GFP signal might also be expected at the Z-discs of the cultured cardiomyocytes. This type of *in vitro* experiments can also be performed with a fluorescently labelled expression construct of *Mospd1*, to see if this Mospd family member shares functional similarities with Mospd3.

In their study, showing that Alp protein can interact with α-actinin to facilitate actin-bundling, Pashmforoush and colleagues found that overexpression of Alp in cultured cardiomyocytes resulted in a marked enhancement of cell cytoarchitecture and sarcomeric organization (Pashmforoush et al., 2001). By studying the subcellular localisation of cytoskeletal proteins, such as actin or α-actinin, in cultured cardiomyocytes which have been transiently transfected with eGFP-Mospd3 compared to untransfected cells, it should be possible to see if Mospd3 is involved in the structural organisation and integration of myofibrils.

A further avenue for studying the function of Mospd proteins *in vitro* is by silencing the expression of *Mospd3* or *Mospd1* through RNAi (RNA interference) in cultured cardiomyocytes (Kasahara and Aoki, 2005) and study the resulting effect on the physiology of the cells. This RNAi-mediated gene expression knockdown has been successfully used in cultured rat cardiomyocytes to study the importance of the desmosomal protein plakophilin-2 in maintaining cell-cell adhesion at the intercalated discs of adjoining cardiomyocytes (Pieperhoff et al., 2008).

In the study of Mospd protein function, RNAi of select desmosomal and adherens junction proteins could be used in conjunction with transient transfection of *eGFP-Mospd3* in cultured cardiomyocytes. This would allow us to determine if and how the loss of these desmosomal and AJ proteins in these cells affects the localisation of the fluorescently labelled Mospd3 protein.

6.2.3 *In vivo* mouse studies

Whilst *in vitro* studies are a key tool to understanding the function of a particular gene, the results gained from these studies generally require verification in the *in vivo* animal model. Gene knockin and knockout techniques provide an invaluable tool to gain insight into the *in vivo* function of a gene. Analysis of the expression pattern of Mospd3 or Mospd1 *in vivo* in the developing and adult mouse model could be facilitated by a targeted knockin of a reporter gene, such as *eGFP* or *β-gal*, and a positive selection marker downstream of the endogenous *Mospd3* or *Mospd1* promoter sequence. The expression pattern of the reporter gene should mirror that of the endogenous gene. As gene targeting allows precise modification of a target locus, problems like the complex reporter gene integration observed in the

gene trapping experiment of *Mospd3* can be circumvented with this method. Also, depending on the placement of the reporter gene cassette and its composition, the expression of the endogenous gene at the targeted locus can either be conserved or disrupted. The second scenario would lead to the constitutive ablation of gene expression from the targeted allele, therefore allowing parallel *in vivo* analysis of a loss-of-function phenotype of the *Mospd* gene of interest in the animal model.

However, for the purpose of gaining more detailed insight into the tissue and developmental stage specific effects of *Mospd3* or *Mospd1* inactivation, we have chosen to generate conditional knockout alleles of these two genes. Once transgenic mouse lines carrying the conditional alleles have been successfully generated, either gene can be inactivated in a tissue and time-dependent manner.

Initial crossing of mice carrying the conditional knockout allele to a ubiquitous *Cre* deleter mouse strain will allow the assessment of complete loss of *Mospd* gene expression on mouse development and phenotype. This study would involve detailed histological analysis to determine which tissues, if any, are affected; as well as close monitoring of the Mendelian ratio of mice born to determine whether loss of expression of either gene results in embryonic or neonatal death. If either gene is, as speculated, involved in cell-cell adhesion complexes, one might expect to see defects in heart structure and possibly even skin blistering. If the loss of *Mospd1* or *Mospd3* expression causes embryonic lethality, or a complex phenotype involving multiple organs or different stages of embryonic development, it may be necessary to delete *Mospd1* or *Mospd3* in a specific tissue or at a defined time point during development. (A schematic showing the conditional knockout mouse lines of *Mospd1* and *Mospd3* –detailed in the remainder of this section - and their breeding, dependent on the observed knockout phenotype, is shown in figure 6.1.)

The requirement of *Mospd* genes specifically in the heart could thus be assessed by mating the conditional knockout (cko) mouse line of *Mospd1* or *Mospd3* to an effecter strain expressing Cre from a cardiac specific promoter, such as MHC-Cre or Nkx2.5-Cre. Similarly, if evidence from the *Mospd1* and *Mospd3* null animals indicates that either gene is involved in the structural integrity of the heart instead of cardiac morphogenesis in the embryo, deletion of either gene in adult animals *via* an

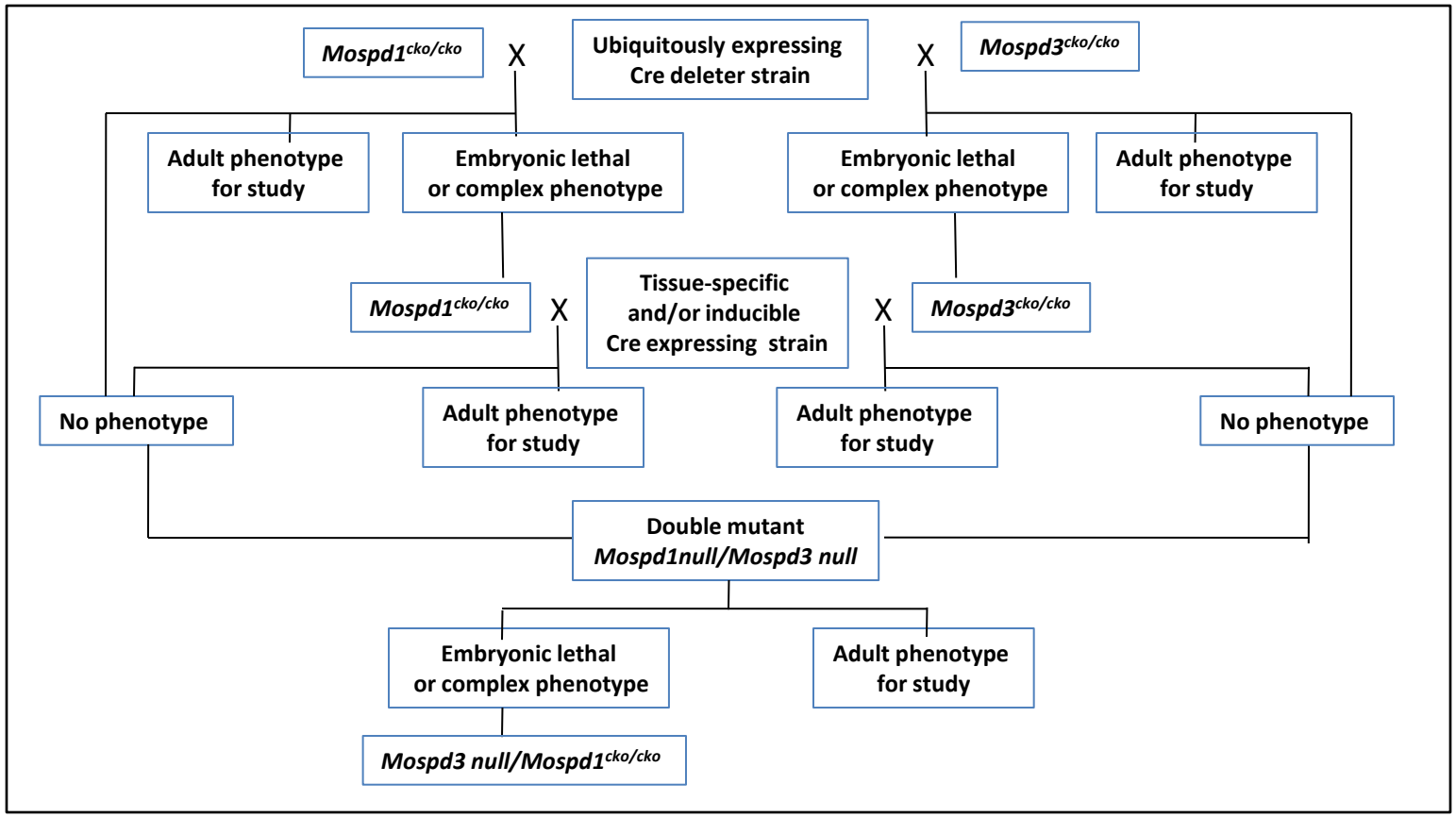


Figure 6.1: Breeding scheme for transgenic mouse lines carrying the conditional knockout allele of *Mospd1* or *Mospd3*. This figure shows mating of the transgenic *Mospd* mouse lines to different Cre expressing strains and to each other. Also indicated are the possible phenotype outcomes of each mating.

inducible Cre can be used to assess this possibility. Inducible deletion can be achieved by mating transgenic animals carrying the conditional knockout allele of *Mospd1* or *Mospd3* to a transgenic mouse line carrying a Cre recombinase transgene which can be induced by either a doxycycline through the tet-"on" system or *via* a tamoxifen or RU486-inducible mutant ligand binding domain system.

Using either of these two systems, *Mospd1* and *Mospd3*can be inactivated in adult animals by administration of either doxycycline or synthetic steroid by injection or through their food (Lewandoski, 2001; Zhang et al., 1996).

It is possible that the deletion of neither gene results in a detectable phenotype or only a mild defect due to genetic redundancy. If this is the case, the generation of a *Mospd1/Mospd3* double mutant should result in a more severe phenotype for further study. If the resulting double mutant animals display a complex phenotype or embryonic lethality, it may be necessary to generate a double mutant carrying a null allele of one of the two *Mospd* genes and a conditional allele of the other, like *Mospd3 null/Mospd1*$^{cko/cko}$. Using a tissue-specific or inducible Cre expressing mouse strain, the conditional allele can then be inactivated to generate a double null allele in a desired tissue or at a specific time point.

The proposed studies of the loss of function of *Mospd1* and *Mospd3* in the animal and *in vitro* model, in addition to the Mospd protein analysis, will help assign a function to the currently uncharacterised *Mospd* gene family.

6.2.4 The zebrafish as a model system for cardiac development and structural integrity of the heart muscle

Evolutionary analysis of *Mospd1* and *Mospd3*, two members of the Mospd gene family in mouse, has shown that *Mospd3* is a mammalian specific gene, whilst orthologs of *Mospd1* were found in a wide range of vertebrate species including chicken, *Xenopus, Fugu* and zebrafish (Pall et al., 2004). The presence of *Mospd1* but not *Mospd3* makes the zebrafish a promising complementary model organism for studying the function of *Mospd* genes in the cardiac system without the complexity of genetic redundancy.

Several aspects of mammalian cardiac development, such as early cardiogenic specification and heart tube formation, appear to be mirrored in the

zebrafish. Genetic determinants of heart tube formation and cardiac looping have also been shown to be largely conserved between zebrafish and mammals (Stainier, 2001). Hence, despite the obvious anatomical differences between the four-chambered mammalian heart and the two-chambered heart of the fish, zebrafish have been successfully used to model aspects of human cardiomyopathies (Gerull et al., 2002; Heuser et al., 2006; Xu et al., 2002).

The zebrafish might prove to be a very useful tool for studying the effect of *Mospd* genes on cardiac development and the structural integrity of the heart as: 1) zebrafish only posses *Mospd1* and not *Mospd3*, avoiding the issue of genetic redundancy; 2) the large number of embryos produced, as well as the use of gene expression knockdown by morpholino injection, enables high throughput gene function study in this model system; and 3) the zebrafish is transparent allowing all organs to be visualised during development, and heart contractility to be measured non-invasively by electrocardiogram methods described by (Milan et al., 2006).

Work on studying the importance of *mospd1* in zebrafish is currently underway in the laboratory. Initial zebrafish *mospd1* morpholino knockdown experiments (carried out with the help of, and now continued by Madina Kara) indicate that expressional down-regulation of this gene in zebrafish embryos leads to cardiac oedema and blood pooling at day 2 to 3 of development. This phenotype may be caused by defects in the structural integrity of the heart and aberrant contractility caused by the loss of desmosomal structures similar to that found in desmocollin-2 ablation in zebrafish (Heuser et al., 2006). As in the mouse model, it will be important to correlate the localisation of the mospd1 protein with desmosomal and adherens junction components to verify our hypothesis of involvement of this protein in cell-cell adhesion at the intercalated disc of cardiomyocytes.

APPENDIX

Primers

Appendix Table 1: Primer sequences used for *Mospd1* and *Mospd3* targeting vector generation by recombineering.

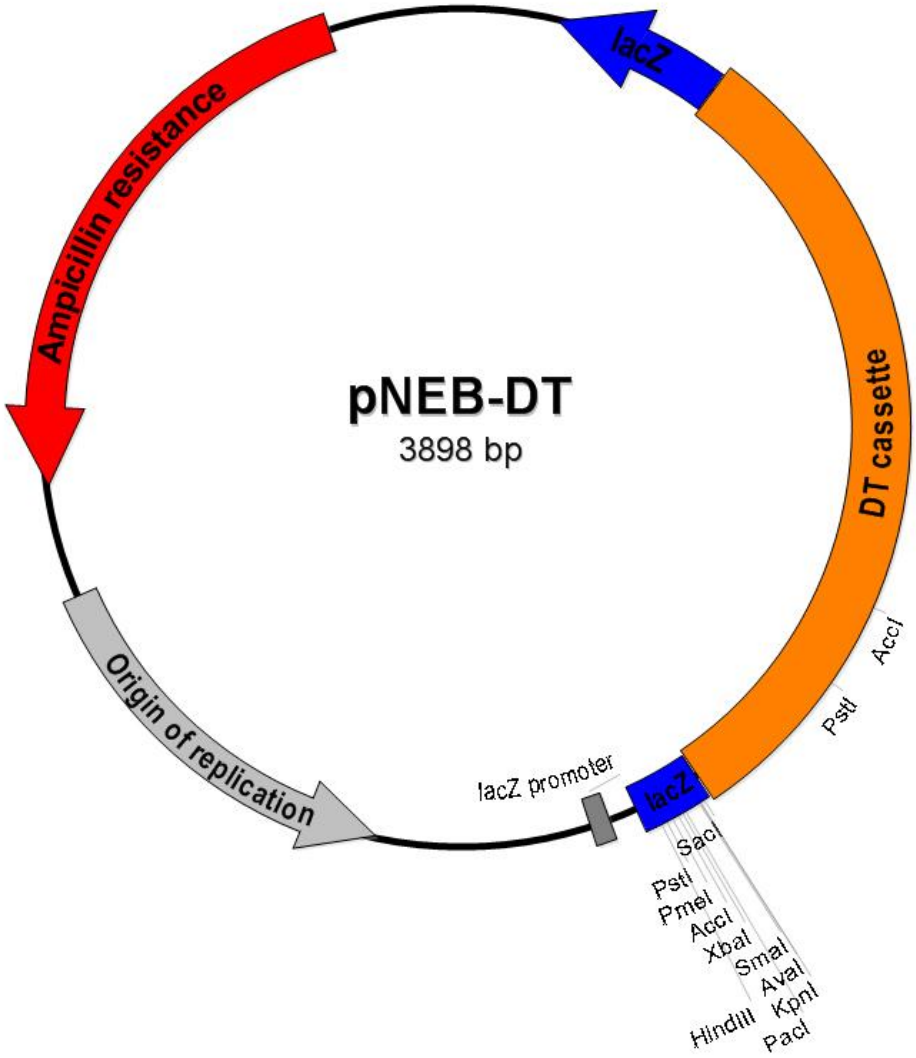
Mospd1	
Mospd1-loxpneo-1_for	5'- ATGATCCTGATTCAATTCCCAGCACCCATGGCAGCA CACAGCTATGTAGAACTATGTAGCTCGACCTGCAGCCA AGCTATCG -3'
Mospd1-loxpneo-2_rev	5'- GCACATGACTTTAATCCAACACTCAGAAGGATTCAG GTACATCTCTTGTGAGTTCCATGACGGCCGCTCTAGAAC TAGTG -3'
Mospd1-FRTneo_p1	5'- GGAGGGTCTAGACTTTCTGCTTAATGCTAGAACTAT GGACTAGGCCTCAGCGATAAGCTTGATATCGAATTC-3
Mospd1-FRTneo_p2	5'-CAAAGTTCTCTCTCACTGAAGAGAAAACGGAGGCCCT GTCAGTGAATATTTCTGCGGCCGCTCTAGAACTAG -3'
Mospd3	
Mospd-3A'-HindIII	5'-CGTCATAAGCTTAGTTTGTATCAGATCTCAGTGTCA -3'
Mospd-3B'-BamHI	5'- GGTGTGGATCCAGGTTGGGATTTAGGTTTAGTG -3'
Mospd-3Y_3-BamHI	5' - GTACGGATCCCTGTTTGTCAGCTGCATC -3'
Mospd-3Z_3AvaI	5'- CAATCCCGAGATGGTTCAGCCACTATTAGC -3'
Mospd3-loxpneo-1_for	5'-CACGTATTGTCAGAGGTCAAATGAGGGCGTTCGATT CCCAGAAACTGGAGTTATAGTGGTACCTCGACCTGCAG CCAAGCTATCG-3'
Mospd3-loxpneo-2_rev	5'-TTAAATAAGAGGGTTAGTGCTGAAGGGTTCGGACCC AGGTTCGCTTCCAGCATAGACATACGGCCGCTCTAGAA CTAGTG -3'
Mospd3-FRTneo_p1	5'-ACTTGAGCACATGTGCAGTCTCTTCTCACCCTTAGTT GTTTTGAGACAGGCGATAAGCTTGATATCGAATTC-3'
Mospd3-FRTneo_p2	5'-TCTGCACCGCCCATCCAGTCTACAAATGAATGTATAT TAGGTTCAATGAGCGGCCGCTCTAGAACTAGTG -3'

Appendix Table 2: Primer sequences used for Southern blot screening of targeted ES cell clones of *Mospd1* and *Mospd3.*

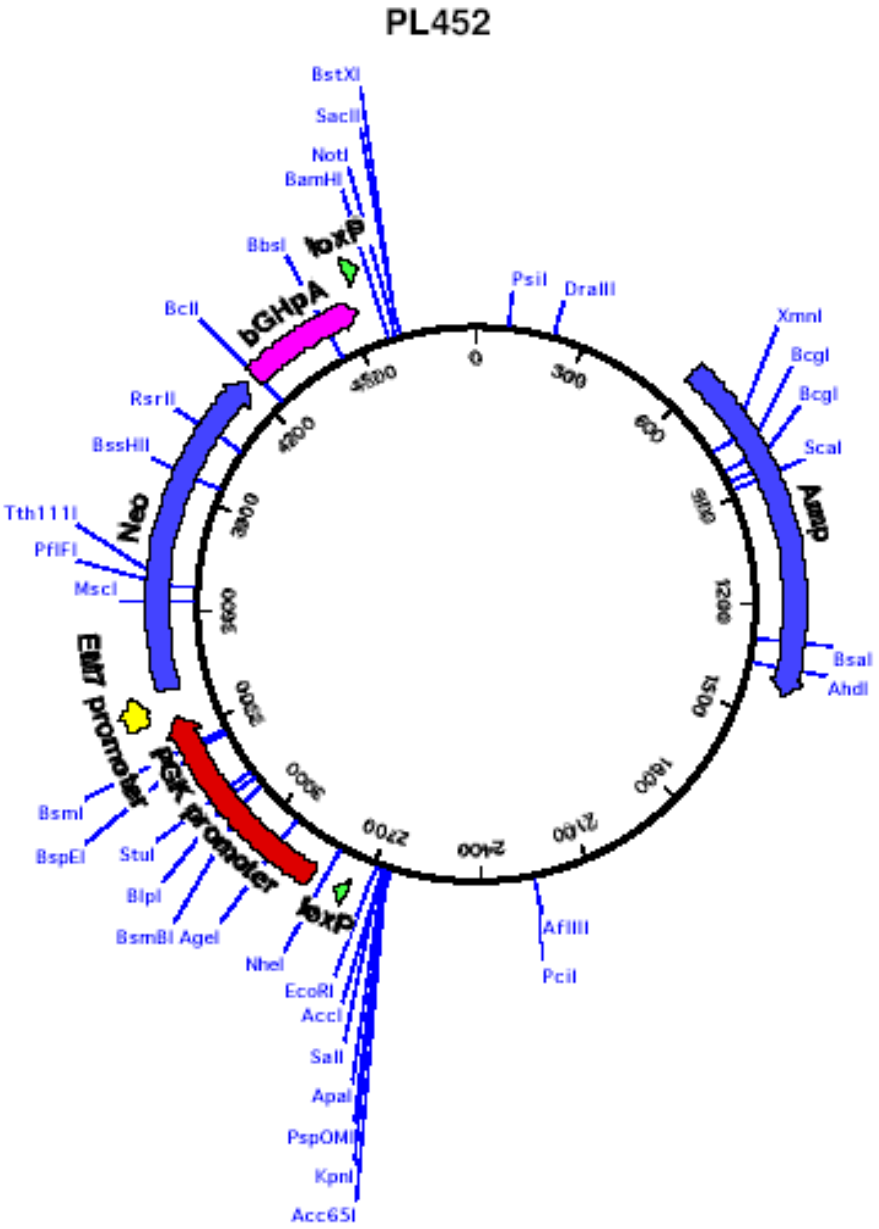
Mospd1	
Mospd1-up1_for	5'- CAGAAGGCAGGAGCAGGAAGAC-3'
Mospd1-up1_rev	5'- CACAGGGCATTAGGGACAGAAA-3'
Mospd1-down3_for	5'-GTGCCAAGGATGCGACTCAG-3'
Mospd1-down3_rev	5'-CCAGAAGGCACAGGCTCAAAT-3'
Mospd3	
Mospd3-up1_for	5'- GGCCCCAAACAACCCTGAC-3'
Mospd3-up1_rev	5'- GATCGCTTCCCCTTTTATTTTTGA-3'
Mospd3-down2_for	5'-TCTGGCTCGGGTGAATGTCTGCT -3'
Mospd3-down2_rev	5'-TTTGTTTTGGGGGAAGGGAGTTGG -3'

Plasmid Maps

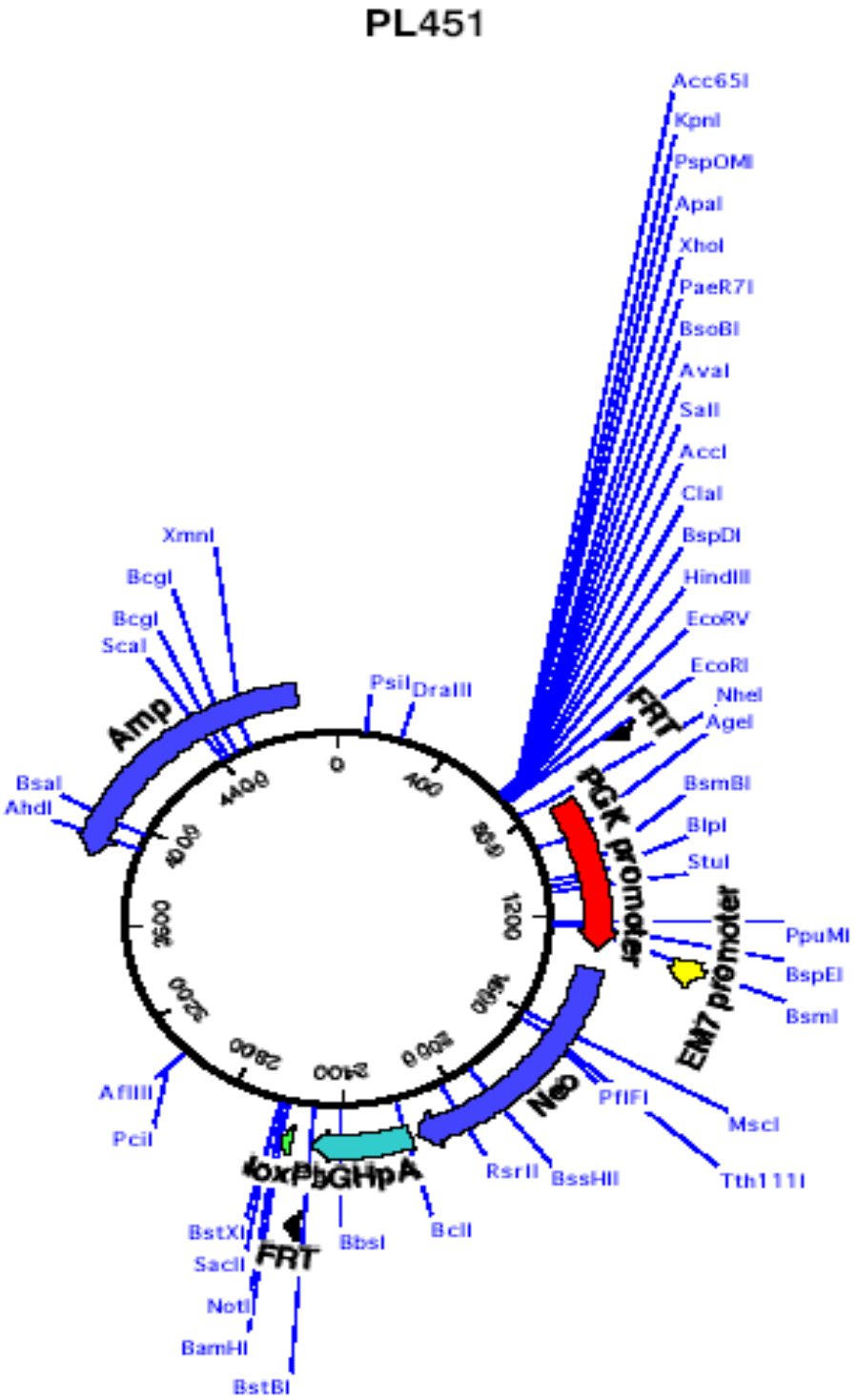
Appendix Figure 1: Plasmid map of pNEB-DT, as constructed by Michele Battle.



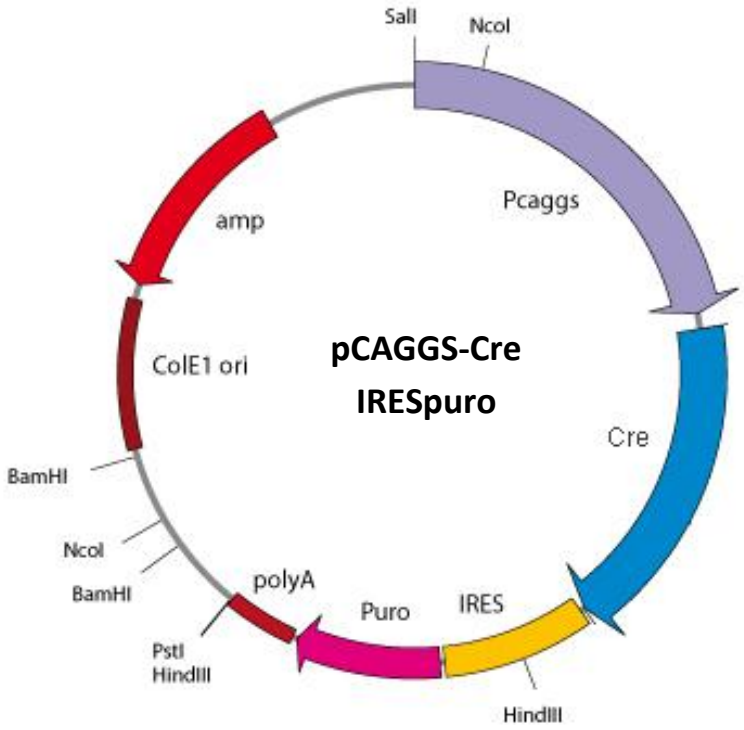
Appendix Figure 2: Plasmid map of PL452, as constructed by Pentao Liu.



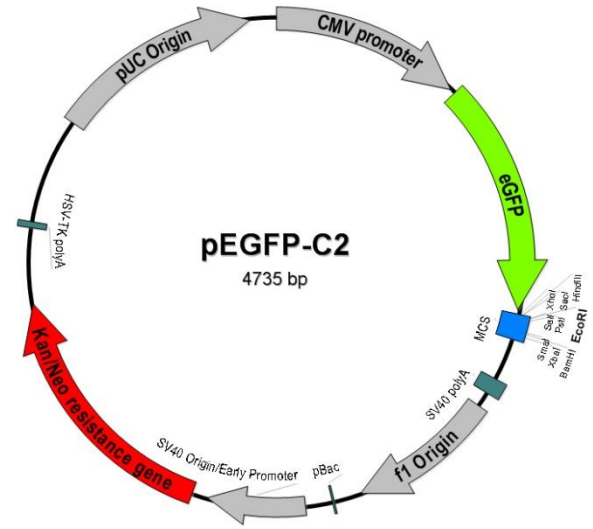
Appendix Figure 3: Plasmid map of PL451 as constructed by Pentao Liu.

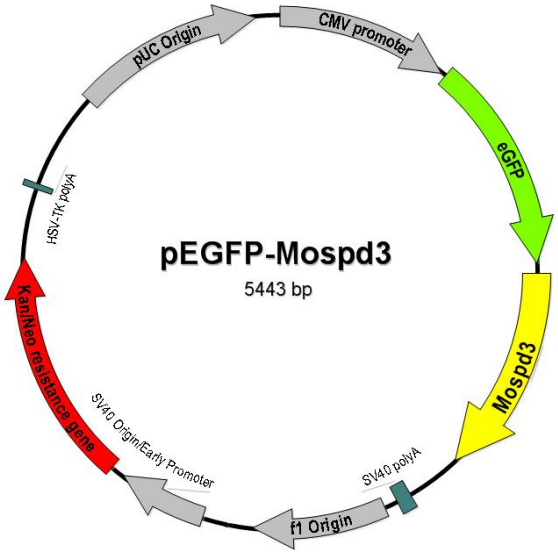


Appendix Figure 4: Plasmid map of pCAGGS-CreIRESpuro.

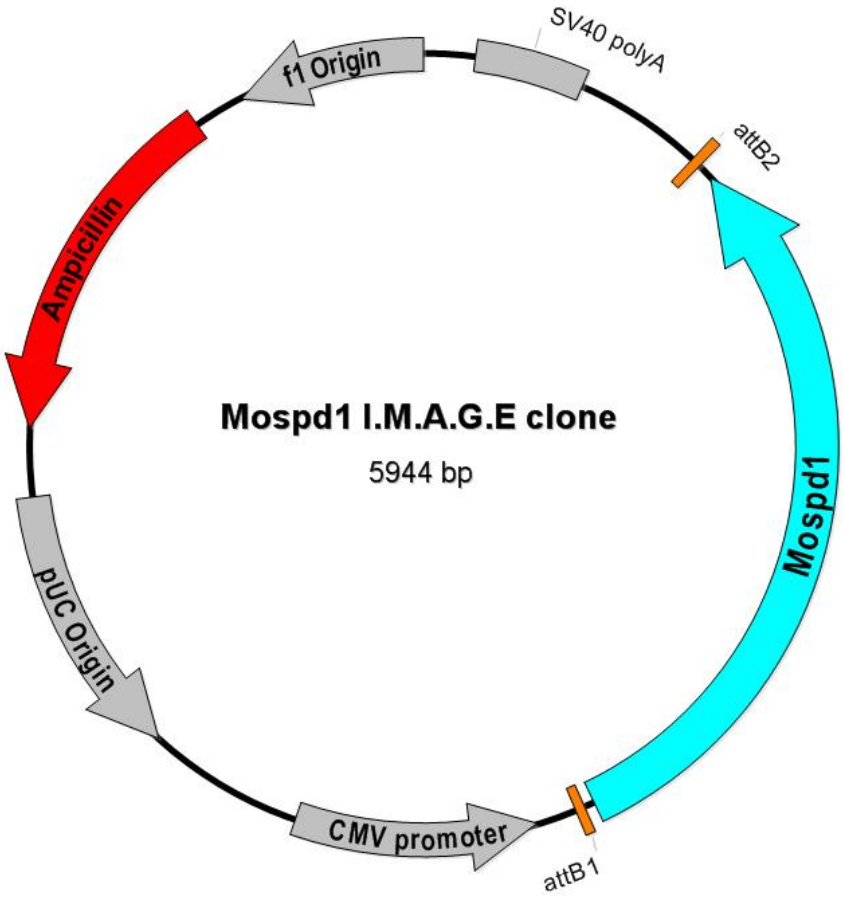


Appendix Figure 5: Plasmid map of the empty pEGFP-C2 vector (upper; Clontech), into which *Mospd3* was subcloned (lower; contructed by Richard Axton).





Appendix Figure 6: Plasmid map of the I.M.A.G.E clone (ID 3962906) containing the *Mospd1* sequence.



REFERENCES

Abu-Issa, R., Smyth, G., Smoak, I., Yamamura, K. and Meyers, E. N. (2002). Fgf8 is required for pharyngeal arch and cardiovascular development in the mouse. *Development* **129**, 4613-25.

Adams, D. J., Quail, M. A., Cox, T., van der Weyden, L., Gorick, B. D., Su, Q., Chan, W. I., Davies, R., Bonfield, J. K., Law, F. et al. (2005). A genome-wide, end-sequenced 129Sv BAC library resource for targeting vector construction. *Genomics* **86**, 753-8.

Albertson, R. C. and Yelick, P. C. (2005). Roles for fgf8 signaling in left-right patterning of the visceral organs and craniofacial skeleton. *Dev Biol* **283**, 310-21.

Araki, K., Araki, M. and Yamamura, K. (2002). Site-directed integration of the cre gene mediated by Cre recombinase using a combination of mutant lox sites. *Nucleic Acids Res* **30**, e103.

Arber, S., Halder, G. and Caroni, P. (1994). Muscle LIM protein, a novel essential regulator of myogenesis, promotes myogenic differentiation. *Cell* **79**, 221-31.

Arber, S., Hunter, J. J., Ross, J., Jr., Hongo, M., Sansig, G., Borg, J., Perriard, J. C., Chien, K. R. and Caroni, P. (1997). MLP-deficient mice exhibit a disruption of cardiac cytoarchitectural organization, dilated cardiomyopathy, and heart failure. *Cell* **88**, 393-403.

Askew, G. R., Doetschman, T. and Lingrel, J. B. (1993). Site-directed point mutations in embryonic stem cells: a gene-targeting tag-and-exchange strategy. *Mol Cell Biol* **13**, 4115-24.

Auerbach, W., Dunmore, J. H., Fairchild-Huntress, V., Fang, Q., Auerbach, A. B., Huszar, D. and Joyner, A. L. (2000). Establishment and chimera analysis of 129/SvEv- and C57BL/6-derived mouse embryonic stem cell lines. *Biotechniques* **29**, 1024-8, 1030, 1032.

Auwerx, J., Avner, P., Baldock, R., Ballabio, A., Balling, R., Barbacid, M., Berns, A., Bradley, A., Brown, S., Carmeliet, P. et al. (2004). The European dimension for the mouse genome mutagenesis program. *Nat Genet* **36**, 925-7.

Awad, M. M., Calkins, H. and Judge, D. P. (2008). Mechanisms of disease: molecular genetics of arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Nat Clin Pract Cardiovasc Med* **5**, 258-67.

Awad, M. M., Dalal, D., Cho, E., Amat-Alarcon, N., James, C., Tichnell, C., Tucker, A., Russell, S. D., Bluemke, D. A., Dietz, H. C. et al. (2006). DSG2 mutations contribute to arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Am J Hum Genet* **79**, 136-42.

Baker, R. K., Haendel, M. A., Swanson, B. J., Shambaugh, J. C., Micales, B. K. and Lyons, G. E. (1997). In vitro preselection of gene-trapped embryonic stem cell

clones for characterizing novel developmentally regulated genes in the mouse. *Dev Biol* **185**, 201-14.

Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F. and Cullin, C. (1993). A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res* **21**, 3329-30.

Bhargava, J., Shashikant, C. S., Carr, J. L., Juan, H., Bentley, K. L. and Ruddle, F. H. (1999). Direct cloning of genomic DNA by recombinogenic targeting method using a yeast-bacterial shuttle vector, pClasper. *Genomics* **62**, 285-8.

Biben, C. and Harvey, R. P. (1997). Homeodomain factor Nkx2-5 controls left/right asymmetric expression of bHLH gene eHand during murine heart development. *Genes Dev* **11**, 1357-69.

Bierkamp, C., McLaughlin, K. J., Schwarz, H., Huber, O. and Kemler, R. (1996). Embryonic heart and skin defects in mice lacking plakoglobin. *Dev Biol* **180**, 780-5.

Blanchard, A., Ohanian, V. and Critchley, D. (1989). The structure and function of alpha-actinin. *J Muscle Res Cell Motil* **10**, 280-9.

Boheler, K. R., Czyz, J., Tweedie, D., Yang, H. T., Anisimov, S. V. and Wobus, A. M. (2002). Differentiation of pluripotent embryonic stem cells into cardiomyocytes. *Circ Res* **91**, 189-201.

Bonaldo, P., Chowdhury, K., Stoykova, A., Torres, M. and Gruss, P. (1998). Efficient gene trap screening for novel developmental genes using IRES beta geo vector and in vitro preselection. *Exp Cell Res* **244**, 125-36.

Borisy, G. G. and Svitkina, T. M. (2000). Actin machinery: pushing the envelope. *Curr Opin Cell Biol* **12**, 104-12.

Bornslaeger, E. A., Corcoran, C. M., Stappenbeck, T. S. and Green, K. J. (1996). Breaking the connection: displacement of the desmosomal plaque protein desmoplakin from cell-cell interfaces disrupts anchorage of intermediate filament bundles and alters intercellular junction assembly. *J Cell Biol* **134**, 985-1001.

Borrmann, C. M., Grund, C., Kuhn, C., Hofmann, I., Pieperhoff, S. and Franke, W. W. (2006). The area composita of adhering junctions connecting heart muscle cells of vertebrates. II. Colocalizations of desmosomal and fascia adhaerens molecules in the intercalated disk. *Eur J Cell Biol* **85**, 469-85.

Bramucci, M. G. and Nagarajan, V. (1996). Direct selection of cloned DNA in Bacillus subtilis based on sucrose-induced lethality. *Appl Environ Microbiol* **62**, 3948-53.

Brook, F. A. and Gardner, R. L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *Proc Natl Acad Sci U S A* **94**, 5709-12.

Buchholz, F., Angrand, P. O. and Stewart, A. F. (1998). Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat Biotechnol* **16**, 657-62.

Bullock, T. L., Roberts, T. M. and Stewart, M. (1996). 2.5 A resolution crystal structure of the motile major sperm protein (MSP) of Ascaris suum. *J Mol Biol* **263**, 284-96.

Burke, D. J. and Ward, S. (1983). Identification of a large multigene family encoding the major sperm protein of Caenorhabditis elegans. *J Mol Biol* **171**, 1-29.

Buttery, L. D., Bourne, S., Xynos, J. D., Wood, H., Hughes, F. J., Hughes, S. P., Episkopou, V. and Polak, J. M. (2001). Differentiation of osteoblasts and in vitro bone formation from murine embryonic stem cells. *Tissue Eng* **7**, 89-99.

Buttery, S. M., Ekman, G. C., Seavy, M., Stewart, M. and Roberts, T. M. (2003). Dissection of the Ascaris sperm motility machinery identifies key proteins involved in major sperm protein-based amoeboid locomotion. *Mol Biol Cell* **14**, 5082-8.

Buxton, R. S. and Magee, A. I. (1992). Structure and interactions of desmosomal and other cadherins. *Semin Cell Biol* **3**, 157-67.

Cai, C. L., Liang, X., Shi, Y., Chu, P. H., Pfaff, S. L., Chen, J. and Evans, S. (2003). Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* **5**, 877-89.

Capdevila, J., Vogan, K. J., Tabin, C. J. and Izpisua Belmonte, J. C. (2000). Mechanisms of left-right determination in vertebrates. *Cell* **101**, 9-21.

Capetanaki, Y. (2000). Desmin cytoskeleton in healthy and failing heart. *Heart Fail Rev* **5**, 203-20.

Carter, D. M. and Radding, C. M. (1971). The role of exonuclease and beta protein of phage lambda in genetic recombination. II. Substrate specificity and the mode of action of lambda exonuclease. *J Biol Chem* **246**, 2502-12.

Chan, W., Costantino, N., Li, R., Lee, S. C., Su, Q., Melvin, D., Court, D. L. and Liu, P. (2007). A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic Acids Res* **35**, e64.

Cheeseman, I. M. and Desai, A. (2005). A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci STKE* **2005**, pl1.

Chen, H., Shi, S., Acosta, L., Li, W., Lu, J., Bao, S., Chen, Z., Yang, Z., Schneider, M. D., Chien, K. R. et al. (2004). BMP10 is essential for maintaining cardiac growth during murine cardiogenesis. *Development* **131**, 2219-31.

Cheng, X., Mihindukulasuriya, K., Den, Z., Kowalczyk, A. P., Calkins, C. C., Ishiko, A., Shimizu, A. and Koch, P. J. (2004). Assessment of splice variant-specific functions of desmocollin 1 in the skin. *Mol Cell Biol* **24**, 154-63.

Clark, A. J., Sandler, S. J., Willis, D. K., Chu, C. C., Blanar, M. A. and Lovett, S. T. (1984). Genes of the RecE and RecF pathways of conjugational recombination in Escherichia coli. *Cold Spring Harb Symp Quant Biol* **49**, 453-62.

Conlon, R. A. and Rossant, J. (1992). Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes in vivo. *Development* **116**, 357-68.

Copeland, N. G., Jenkins, N. A. and Court, D. L. (2001). Recombineering: a powerful new tool for mouse functional genomics. *Nat Rev Genet* **2**, 769-79.

Cosloy, S. D. and Oishi, M. (1973). Genetic transformation in Escherichia coli K12. *Proc Natl Acad Sci U S A* **70**, 84-7.

Datsenko, K. A. and Wanner, B. L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-5.

Deng, C. and Capecchi, M. R. (1992). Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Mol Cell Biol* **12**, 3365-71.

Deng, J. M. and Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Transgenic Res* **4**, 264-9.

Detloff, P. J., Lewis, J., John, S. W., Shehee, W. R., Langenbach, R., Maeda, N. and Smithies, O. (1994). Deletion and replacement of the mouse adult beta-globin genes by a "plug and socket" repeated targeting strategy. *Mol Cell Biol* **14**, 6936-43.

Diaz, V., Servert, P., Prieto, I., Gonzalez, M. A., Martinez, A. C., Alonso, J. C. and Bernad, A. (2001). New insights into host factor requirements for prokaryotic beta-recombinase-mediated reactions in mammalian cells. *J Biol Chem* **276**, 16257-64.

Djinovic-Carugo, K., Young, P., Gautel, M. and Saraste, M. (1999). Structure of the alpha-actinin rod: molecular basis for cross-linking of actin filaments. *Cell* **98**, 537-46.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. and Kemler, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J Embryol Exp Morphol* **87**, 27-45.

Ehler, E., Horowits, R., Zuppinger, C., Price, R. L., Perriard, E., Leu, M., Caroni, P., Sussman, M., Eppenberger, H. M. and Perriard, J. C. (2001). Alterations at the intercalated disk associated with the absence of muscle LIM protein. *J Cell Biol* **153**, 763-72.

Ellis, H. M., Yu, D., DiTizio, T. and Court, D. L. (2001). High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98**, 6742-6.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-6.

Farrell, M., Waldo, K., Li, Y. X. and Kirby, M. L. (1999). A novel role for cardiac neural crest in heart development. *Trends Cardiovasc Med* **9**, 214-20.

Fassler, R., Rohwedel, J., Maltsev, V., Bloch, W., Lentini, S., Guan, K., Gullberg, D., Hescheler, J., Addicks, K. and Wobus, A. M. (1996). Differentiation and integrity of cardiac muscle cells are impaired in the absence of beta 1 integrin. *J Cell Sci* **109 (Pt 13)**, 2989-99.

Firulli, A. B., McFadden, D. G., Lin, Q., Srivastava, D. and Olson, E. N. (1998). Heart and extra-embryonic mesodermal defects in mouse embryos lacking the bHLH transcription factor Hand1. *Nat Genet* **18**, 266-70.

Forrester, L. M., Nagy, A., Sam, M., Watt, A., Stevenson, L., Bernstein, A., Joyner, A. L. and Wurst, W. (1996). An induction gene trap screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro. *Proc Natl Acad Sci U S A* **93**, 1677-82.

Fraichard, A., Chassande, O., Bilbaut, G., Dehay, C., Savatier, P. and Samarut, J. (1995). In vitro differentiation of embryonic stem cells into glial cells and functional neurons. *J Cell Sci* **108 (Pt 10)**, 3181-8.

Franke, W. W., Borrmann, C. M., Grund, C. and Pieperhoff, S. (2006). The area composita of adhering junctions connecting heart muscle cells of vertebrates. I. Molecular definition in intercalated disks of cardiomyocytes by immunoelectron microscopy of desmosomal proteins. *Eur J Cell Biol* **85**, 69-82.

Friedrich, G. and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev* **5**, 1513-23.

Frohman, M. A. and Martin, G. R. (1989). Cut, paste, and save: new approaches to altering specific genes in mice. *Cell* **56**, 145-7.

Gallicano, G. I., Bauer, C. and Fuchs, E. (2001). Rescuing desmoplakin function in extra-embryonic ectoderm reveals the importance of this protein in embryonic heart, neuroepithelium, skin and vasculature. *Development* **128**, 929-41.

Gallicano, G. I., Kouklis, P., Bauer, C., Yin, M., Vasioukhin, V., Degenstein, L. and Fuchs, E. (1998). Desmoplakin is required early in development for assembly of desmosomes and cytoskeletal linkage. *J Cell Biol* **143**, 2009-22.

Garcia-Gras, E., Lombardi, R., Giocondo, M. J., Willerson, J. T., Schneider, M. D., Khoury, D. S. and Marian, A. J. (2006). Suppression of canonical Wnt/beta-catenin signaling by nuclear plakoglobin recapitulates phenotype of arrhythmogenic right ventricular cardiomyopathy. *J Clin Invest* **116**, 2012-21.

Garratt, A. N., Ozcelik, C. and Birchmeier, C. (2003). ErbB2 pathways in heart and neural diseases. *Trends Cardiovasc Med* **13**, 80-6.

Gassmann, M., Casagranda, F., Orioli, D., Simon, H., Lai, C., Klein, R. and Lemke, G. (1995). Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature* **378**, 390-4.

Gerull, B., Gramlich, M., Atherton, J., McNabb, M., Trombitas, K., Sasse-Klaassen, S., Seidman, J. G., Seidman, C., Granzier, H., Labeit, S. et al. (2002). Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet* **30**, 201-4.

Gerull, B., Heuser, A., Wichter, T., Paul, M., Basson, C. T., McDermott, D. A., Lerman, B. B., Markowitz, S. M., Ellinor, P. T., MacRae, C. A. et al. (2004). Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet* **36**, 1162-4.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W. and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766-9.

Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C. (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-5.

Gottlieb, P. D., Pierce, S. A., Sims, R. J., Yamagishi, H., Weihe, E. K., Harriss, J. V., Maika, S. D., Kuziel, W. A., King, H. L., Olson, E. N. et al. (2002). Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis. *Nat Genet* **31**, 25-32.

Grant, R. P., Buttery, S. M., Ekman, G. C., Roberts, T. M. and Stewart, M. (2005). Structure of MFP2 and its function in enhancing MSP polymerization in Ascaris sperm amoeboid motility. *J Mol Biol* **347**, 583-95.

Green, K. J. and Gaudry, C. A. (2000). Are desmosomes more than tethers for intermediate filaments? *Nat Rev Mol Cell Biol* **1**, 208-16.

Grossmann, K. S., Grund, C., Huelsken, J., Behrend, M., Erdmann, B., Franke, W. W. and Birchmeier, W. (2004). Requirement of plakophilin 2 for heart morphogenesis and cardiac junction formation. *J Cell Biol* **167**, 149-60.

Gu, H., Marth, J. D., Orban, P. C., Mossmann, H. and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **265**, 103-6.

Guan, K., Czyz, J., Furst, D. O. and Wobus, A. M. (2001). Expression and cellular distribution of alpha(v)integrins in beta(1)integrin-deficient embryonic stem cell-derived cardiac cells. *J Mol Cell Cardiol* **33**, 521-32.

Harlow, E. and Lane, D. (1988). Antibodies : a laboratory manual. New York: Cold Spring Harbor Laboratory.

Harris, B. Z. and Lim, W. A. (2001). Mechanism and role of PDZ domains in signaling complex assembly. *J Cell Sci* **114**, 3219-31.

Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991a). Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells. *Nature* **350**, 243-6.

Hasty, P., Rivera-Perez, J. and Bradley, A. (1991b). The length of homology required for gene targeting in embryonic stem cells. *Mol Cell Biol* **11**, 5586-91.

Henderson, J. R., Pomies, P., Auffray, C. and Beckerle, M. C. (2003). ALP and MLP distribution during myofibrillogenesis in cultured cardiomyocytes. *Cell Motil Cytoskeleton* **54**, 254-65.

Herrick, T. M. and Cooper, J. A. (2002). A hypomorphic allele of dab1 reveals regional differences in reelin-Dab1 signaling during brain development. *Development* **129**, 787-96.

Hescheler, J., Fleischmann, B. K., Lentini, S., Maltsev, V. A., Rohwedel, J., Wobus, A. M. and Addicks, K. (1997). Embryonic stem cells: a model to study structural and functional properties in cardiomyogenesis. *Cardiovasc Res* **36**, 149-62.

Heuser, A., Plovie, E. R., Ellinor, P. T., Grossmann, K. S., Shin, J. T., Wichter, T., Basson, C. T., Lerman, B. B., Sasse-Klaassen, S., Thierfelder, L. et al. (2006). Mutant desmocollin-2 causes arrhythmogenic right ventricular cardiomyopathy. *Am J Hum Genet* **79**, 1081-8.

Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292-5.

Huen, A. C., Park, J. K., Godsel, L. M., Chen, X., Bannon, L. J., Amargo, E. V., Hudson, T. Y., Mongiu, A. K., Leigh, I. M., Kelsell, D. P. et al. (2002). Intermediate filament-membrane attachments function synergistically with actin-dependent contacts to regulate intercellular adhesive strength. *J Cell Biol* **159**, 1005-17.

Ilagan, R., Abu-Issa, R., Brown, D., Yang, Y. P., Jiao, K., Schwartz, R. J., Klingensmith, J. and Meyers, E. N. (2006). Fgf8 is required for anterior heart field development. *Development* **133**, 2435-45.

Italiano, J. E., Jr., Roberts, T. M., Stewart, M. and Fontana, C. A. (1996). Reconstitution in vitro of the motile apparatus from the amoeboid sperm of Ascaris shows that filament assembly and bundling move membranes. *Cell* **84**, 105-14.

Italiano, J. E., Jr., Stewart, M. and Roberts, T. M. (2001). How the assembly dynamics of the nematode major sperm protein generate amoeboid cell motility. *Int Rev Cytol* **202**, 1-34.

Izsvak, Z. and Ivics, Z. (2005). Sleeping Beauty hits them all: transposon-mediated saturation mutagenesis in the mouse germline. *Nat Methods* **2**, 735-6.

Jeannotte, L., Ruiz, J. C. and Robertson, E. J. (1991). Low level of Hox1.3 gene expression does not preclude the use of promoterless vectors to generate a targeted gene disruption. off. *Mol Cell Biol* **11**, 5578-85.

Joyner, A. L. (1999). Gene targeting : a practical approach. Oxford ; New York: Oxford University Press.

Kagiwada, S., Hosaka, K., Murata, M., Nikawa, J. and Takatsuki, A. (1998). The Saccharomyces cerevisiae SCS2 gene product, a homolog of a synaptobrevin-associated protein, is an integral membrane protein of the endoplasmic reticulum and is required for inositol metabolism. *J Bacteriol* **180**, 1700-8.

Kasahara, H. and Aoki, H. (2005). Gene silencing using adenoviral RNAi vector in vascular smooth muscle cells and cardiomyocytes. *Methods Mol Med* **112**, 155-72.

Kellendonk, C., Tronche, F., Monaghan, A. P., Angrand, P. O., Stewart, F. and Schutz, G. (1996). Regulation of Cre recombinase activity by the synthetic steroid RU 486. *Nucleic Acids Res* **24**, 1404-11.

Keller, G. M. (1995). In vitro differentiation of embryonic stem cells. *Curr Opin Cell Biol* **7**, 862-9.

Kelly, R. G., Brown, N. A. and Buckingham, M. E. (2001). The arterial pole of the mouse heart forms from Fgf10-expressing cells in pharyngeal mesoderm. *Dev Cell* **1**, 435-40.

Khurana, T., Khurana, B. and Noegel, A. A. (2002). LIM proteins: association with the actin cytoskeleton. *Protoplasma* **219**, 1-12.

King, K. L., Essig, J., Roberts, T. M. and Moerland, T. S. (1994a). Regulation of the Ascaris major sperm protein (MSP) cytoskeleton by intracellular pH. *Cell Motil Cytoskeleton* **27**, 193-205.

King, K. L., Stewart, M. and Roberts, T. M. (1994b). Supramolecular assemblies of the Ascaris suum major sperm protein (MSP) associated with amoeboid cell motility. *J Cell Sci* **107 (Pt 10)**, 2941-9.

Kirchhof, P., Fabritz, L., Zwiener, M., Witt, H., Schafers, M., Zellerhoff, S., Paul, M., Athai, T., Hiller, K. H., Baba, H. A. et al. (2006). Age- and training-dependent development of arrhythmogenic right ventricular cardiomyopathy in heterozygous plakoglobin-deficient mice. *Circulation* **114**, 1799-806.

Kistner, A., Gossen, M., Zimmermann, F., Jerecic, J., Ullmer, C., Lubbert, H. and Bujard, H. (1996). Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. *Proc Natl Acad Sci U S A* **93**, 10933-8.

Kluppel, M., Vallis, K. A. and Wrana, J. L. (2002). A high-throughput induction gene trap approach defines C4ST as a target of BMP signaling. *Mech Dev* **118**, 77-89.

Klymkowsky, M. W. (2005). beta-catenin and its regulatory network. *Hum Pathol* **36**, 225-7.

Knoll, R., Hoshijima, M., Hoffman, H. M., Person, V., Lorenzen-Schmidt, I., Bang, M. L., Hayashi, T., Shiga, N., Yasukawa, H., Schaper, W. et al. (2002). The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell* **111**, 943-55.

Kostetskii, I., Li, J., Xiong, Y., Zhou, R., Ferrari, V. A., Patel, V. V., Molkentin, J. D. and Radice, G. L. (2005). Induced deletion of the N-cadherin gene in the heart leads to dissolution of the intercalated disc structure. *Circ Res* **96**, 346-54.

Kramer, J., Hegert, C., Guan, K., Wobus, A. M., Muller, P. K. and Rohwedel, J. (2000). Embryonic stem cell-derived chondrogenic differentiation in vitro: activation by BMP-2 and BMP-4. *Mech Dev* **92**, 193-205.

Kuo, C. T., Morrisey, E. E., Anandappa, R., Sigrist, K., Lu, M. M., Parmacek, M. S., Soudais, C. and Leiden, J. M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* **11**, 1048-60.

Laurent, F., Labesse, G. and de Wit, P. (2000). Molecular cloning and partial characterization of a plant VAP33 homologue with a major sperm protein domain. *Biochem Biophys Res Commun* **270**, 286-92.

LeClaire, L. L., 3rd, Stewart, M. and Roberts, T. M. (2003). A 48 kDa integral membrane phosphoprotein orchestrates the cytoskeletal dynamics that generate amoeboid cell motility in Ascaris sperm. *J Cell Sci* **116**, 2655-63.

Lee, E. C., Yu, D., Martinez de Velasco, J., Tessarollo, L., Swing, D. A., Court, D. L., Jenkins, N. A. and Copeland, N. G. (2001). A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73**, 56-65.

Lewandoski, M. (2001). Conditional control of gene expression in the mouse. *Nat Rev Genet* **2**, 743-55.

Li, Z., Colucci-Guyon, E., Pincon-Raymond, M., Mericskay, M., Pournin, S., Paulin, D. and Babinet, C. (1996). Cardiovascular lesions and skeletal myopathy in mice lacking desmin. *Dev Biol* **175**, 362-6.

Li, Z., Mericskay, M., Agbulut, O., Butler-Browne, G., Carlsson, L., Thornell, L. E., Babinet, C. and Paulin, D. (1997). Desmin is essential for the tensile strength and integrity of myofibrils but not for myogenic commitment, differentiation, and fusion of skeletal muscle. *J Cell Biol* **139**, 129-44.

Lin, Q., Schwarz, J., Bucana, C. and Olson, E. N. (1997). Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276**, 1404-7.

Linask, K. K. (1992). N-cadherin localization in early heart development and polar expression of Na+,K(+)-ATPase, and integrin during pericardial coelom formation and epithelialization of the differentiating myocardium. *Dev Biol* **151**, 213-24.

Liu, P., Jenkins, N. A. and Copeland, N. G. (2003). A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res* **13**, 476-84.

Liu, X., Wu, H., Loring, J., Hormuzdi, S., Disteche, C. M., Bornstein, P. and Jaenisch, R. (1997). Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Dev Dyn* **209**, 85-91.

Lloyd, R. G. and Buckman, C. (1985). Identification and genetic analysis of sbcC mutations in commonly used recBC sbcB strains of Escherichia coli K-12. *J Bacteriol* **164**, 836-44.

Longo, L., Bygrave, A., Grosveld, F. G. and Pandolfi, P. P. (1997). The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. *Transgenic Res* **6**, 321-8.

Lorenzen-Schmidt, I., McCulloch, A. D. and Omens, J. H. (2005). Deficiency of actinin-associated LIM protein alters regional right ventricular function and hypertrophic remodeling. *Ann Biomed Eng* **33**, 888-96.

MacRae, C. A., Birchmeier, W. and Thierfelder, L. (2006). Arrhythmogenic right ventricular cardiomyopathy: moving toward mechanism. *J Clin Invest* **116**, 1825-8.

Maltsev, V. A., Rohwedel, J., Hescheler, J. and Wobus, A. M. (1993). Embryonic stem cells differentiate in vitro into cardiomyocytes representing sinusnodal, atrial and ventricular cell types. *Mech Dev* **44**, 41-50.

Mansour, S. L., Thomas, K. R. and Capecchi, M. R. (1988). Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-52.

Marshall, H., Nonchev, S., Sham, M. H., Muchamore, I., Lumsden, A. and Krumlauf, R. (1992). Retinoic acid alters hindbrain Hox code and induces transformation of rhombomeres 2/3 into a 4/5 identity. *Nature* **360**, 737-41.

Martin, G. R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**, 7634-8.

Marvin, M. J., Di Rocco, G., Gardiner, A., Bush, S. M. and Lassar, A. B. (2001). Inhibition of Wnt activity induces heart formation from posterior mesoderm. *Genes Dev* **15**, 316-27.

McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J. and Forrester, L. (1998). Gene trap integrations expressed in the developing heart: insertion site affects splicing of the PT1-ATG vector. *Dev Dyn* **212**, 267-76.

McFadden, D. G., Barbosa, A. C., Richardson, J. A., Schneider, M. D., Srivastava, D. and Olson, E. N. (2005). The Hand1 and Hand2 transcription factors regulate expansion of the embryonic cardiac ventricles in a gene dosage-dependent manner. *Development* **132**, 189-201.

McKoy, G., Protonotarios, N., Crosby, A., Tsatsopoulou, A., Anastasakis, A., Coonar, A., Norman, M., Baboonian, C., Jeffery, S. and McKenna, W. J. (2000). Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet* **355**, 2119-24.

McMahon, A. P. and Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**, 1073-85.

Metzger, D., Clifford, J., Chiba, H. and Chambon, P. (1995). Conditional site-specific recombination in mammalian cells using a ligand-dependent chimeric Cre recombinase. *Proc Natl Acad Sci U S A* **92**, 6991-5.

Meyer, D. and Birchmeier, C. (1995). Multiple essential functions of neuregulin in development. *Nature* **378**, 386-90.

Meyers, E. N., Lewandoski, M. and Martin, G. R. (1998). An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat Genet* **18**, 136-41.

Milan, D. J., Jones, I. L., Ellinor, P. T. and MacRae, C. A. (2006). In vivo recording of adult zebrafish electrocardiogram and assessment of drug-induced QT prolongation. *Am J Physiol Heart Circ Physiol* **291**, H269-73.

Miller, M. A., Nguyen, V. Q., Lee, M. H., Kosinski, M., Schedl, T., Caprioli, R. M. and Greenstein, D. (2001). A sperm cytoskeletal protein that signals oocyte meiotic maturation and ovulation. *Science* **291**, 2144-7.

Milner, D. J., Taffet, G. E., Wang, X., Pham, T., Tamura, T., Hartley, C., Gerdes, A. M. and Capetanaki, Y. (1999). The absence of desmin leads to cardiomyocyte hypertrophy and cardiac dilation with compromised systolic function. *J Mol Cell Cardiol* **31**, 2063-76.

Milner, D. J., Weitzer, G., Tran, D., Bradley, A. and Capetanaki, Y. (1996). Disruption of muscle architecture and myocardial degeneration in mice lacking desmin. *J Cell Biol* **134**, 1255-70.

Mitchell, K. J., Pinson, K. I., Kelly, O. G., Brennan, J., Zupicich, J., Scherz, P., Leighton, P. A., Goodrich, L. V., Lu, X., Avery, B. J. et al. (2001). Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat Genet* **28**, 241-9.

Mitne-Neto, M., Ramos, C. R., Pimenta, D. C., Luz, J. S., Nishimura, A. L., Gonzales, F. A., Oliveira, C. C. and Zatz, M. (2007). A mutation in human VAP-B--MSP domain, present in ALS patients, affects the interaction with other cellular proteins. *Protein Expr Purif* **55**, 139-46.

Mjaatvedt, C. H., Nakaoka, T., Moreno-Rodriguez, R., Norris, R. A., Kern, M. J., Eisenberg, C. A., Turner, D. and Markwald, R. R. (2001). The outflow tract of the heart is recruited from a novel heart-forming field. *Dev Biol* **238**, 97-109.

Mohapatra, B., Jimenez, S., Lin, J. H., Bowles, K. R., Coveler, K. J., Marx, J. G., Chrisco, M. A., Murphy, R. T., Lurie, P. R., Schwartz, R. J. et al. (2003). Mutations in the muscle LIM protein and alpha-actinin-2 genes in dilated cardiomyopathy and endocardial fibroelastosis. *Mol Genet Metab* **80**, 207-15.

Molkentin, J. D., Lin, Q., Duncan, S. A. and Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11**, 1061-72.

Muyrers, J. P., Zhang, Y., Benes, V., Testa, G., Ansorge, W. and Stewart, A. F. (2000). Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep* **1**, 239-43.

Muyrers, J. P., Zhang, Y., Testa, G. and Stewart, A. F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* **27**, 1555-7.

Nagy, A. (2000). Cre recombinase: the universal reagent for genome tailoring. *Genesis* **26**, 99-109.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J. C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci U S A* **90**, 8424-8.

Nakano, T., Kodama, H. and Honjo, T. (1994). Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098-101.

Narita, N., Bielinska, M. and Wilson, D. B. (1997). Cardiomyocyte differentiation by GATA-4-deficient embryonic stem cells. *Development* **124**, 3755-64.

Nelson, G. A., Roberts, T. M. and Ward, S. (1982). Caenorhabditis elegans spermatozoan locomotion: amoeboid movement with almost no actin. *J Cell Biol* **92**, 121-31.

Nelson, P. N., Reynolds, G. M., Waldron, E. E., Ward, E., Giannopoulos, K. and Murray, P. G. (2000). Monoclonal antibodies. *Mol Pathol* **53**, 111-7.

Nichols, J., Evans, E. P. and Smith, A. G. (1990). Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341-8.

Nishikawa, S. I., Nishikawa, S., Hirashima, M., Matsuyoshi, N. and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. *Development* **125**, 1747-57.

Nishimura, A. L., Mitne-Neto, M., Silva, H. C., Richieri-Costa, A., Middleton, S., Cascio, D., Kok, F., Oliveira, J. R., Gillingwater, T., Webb, J. et al. (2004). A mutation in the vesicle-trafficking protein VAPB causes late-onset spinal muscular atrophy and amyotrophic lateral sclerosis. *Am J Hum Genet* **75**, 822-31.

Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S. and Yamamura, K. (1993). An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J Biochem* **113**, 343-9.

Norgett, E. E., Hatsell, S. J., Carvajal-Huerta, L., Cabezas, J. C., Common, J., Purkis, P. E., Whittock, N., Leigh, I. M., Stevens, H. P. and Kelsell, D. P. (2000). Recessive mutation in desmoplakin disrupts desmoplakin-intermediate filament interactions and causes dilated cardiomyopathy, woolly hair and keratoderma. *Hum Mol Genet* **9**, 2761-6.

O'Connor, M., Peifer, M. and Bender, W. (1989). Construction of large DNA segments in Escherichia coli. *Science* **244**, 1307-12.

O'Gorman, S., Fox, D. T. and Wahl, G. M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-5.

Ohtsuka, M., Ishii, K., Kikuti, Y. Y., Warita, T., Suzuki, D., Sato, M., Kimura, M. and Inoko, H. (2006). Construction of mouse 129/Ola BAC library for targeting experiments using E14 embryonic stem cells. *Genes Genet Syst* **81**, 143-6.

Olkkonen, V. M. and Ikonen, E. (2006). When intracellular logistics fails--genetic defects in membrane trafficking. *J Cell Sci* **119**, 5031-45.

Ornitz, D. M., Moreadith, R. W. and Leder, P. (1991). Binary system for regulating transgene expression in mice: targeting int-2 gene expression with yeast GAL4/UAS control elements. *Proc Natl Acad Sci U S A* **88**, 698-702.

Otey, C. A. and Carpen, O. (2004). Alpha-actinin revisited: a fresh look at an old player. *Cell Motil Cytoskeleton* **58**, 104-11.

Pall, G. S. (1998). PhD Thesis: Characterisation of gene-trap integrations expressed during mouse heart development Edinburgh: University of Edinburgh Library.

Pall, G. S., Wallis, J., Axton, R., Brownstein, D. G., Gautier, P., Buerger, K., Mulford, C., Mullins, J. J. and Forrester, L. M. (2004). A novel transmembrane MSP-containing protein that plays a role in right ventricle development. *Genomics* **84**, 1051-9.

Pashmforoush, M., Pomies, P., Peterson, K. L., Kubalak, S., Ross, J., Jr., Hefti, A., Aebi, U., Beckerle, M. C. and Chien, K. R. (2001). Adult mice deficient in actinin-associated LIM-domain protein reveal a developmental pathway for right ventricular cardiomyopathy. *Nat Med* **7**, 591-7.

Pennetta, G., Hiesinger, P. R., Fabian-Fine, R., Meinertzhagen, I. A. and Bellen, H. J. (2002). Drosophila VAP-33A directs bouton formation at neuromuscular junctions in a dosage-dependent manner. *Neuron* **35**, 291-306.

Perriard, J. C., Hirschy, A. and Ehler, E. (2003). Dilated cardiomyopathy: a disease of the intercalated disc? *Trends Cardiovasc Med* **13**, 30-8.

Phan, D., Rasmussen, T. L., Nakagawa, O., McAnally, J., Gottlieb, P. D., Tucker, P. W., Richardson, J. A., Bassel-Duby, R. and Olson, E. N. (2005). BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart. *Development* **132**, 2669-78.

Pieperhoff, S., Schumacher, H. and Franke, W. W. (2008). The area composita of adhering junctions connecting heart muscle cells of vertebrates. V. The importance of plakophilin-2 demonstrated by small interference RNA-mediated knockdown in cultured rat cardiomyocytes. *Eur J Cell Biol* **87**, 399-411.

Pilichou, K., Nava, A., Basso, C., Beffagna, G., Bauce, B., Lorenzon, A., Frigo, G., Vettori, A., Valente, M., Towbin, J. et al. (2006). Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy. *Circulation* **113**, 1171-9.

Pinto, V. B., Prasad, S., Yewdell, J., Bennink, J. and Hughes, S. H. (2000). Restricting expression prolongs expression of foreign genes introduced into animals by retroviruses. *J Virol* **74**, 10202-6.

Pollard, T. D. and Borisy, G. G. (2003). Cellular motility driven by assembly and disassembly of actin filaments. *Cell* **112**, 453-65.

Pomies, P., Macalma, T. and Beckerle, M. C. (1999). Purification and characterization of an alpha-actinin-binding PDZ-LIM protein that is up-regulated during muscle differentiation. *J Biol Chem* **274**, 29242-50.

Pomies, P., Pashmforoush, M., Vegezzi, C., Chien, K. R., Auffray, C. and Beckerle, M. C. (2007). The cytoskeleton-associated PDZ-LIM protein, ALP, acts on serum response factor activity to regulate muscle differentiation. *Mol Biol Cell* **18**, 1723-33.

Poser, I., Sarov, M., Hutchins, J. R., Heriche, J. K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W. et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods* **5**, 409-15.

Poteete, A. R. (2001). What makes the bacteriophage lambda Red system useful for genetic engineering: molecular mechanism and biological function. *FEMS Microbiol Lett* **201**, 9-14.

Protonotarios, N., Tsatsopoulou, A., Anastasakis, A., Sevdalis, E., McKoy, G., Stratos, K., Gatzoulis, K., Tentolouris, K., Spiliopoulou, C., Panagiotakos, D. et al. (2001). Genotype-phenotype assessment in autosomal recessive arrhythmogenic right ventricular cardiomyopathy (Naxos disease) caused by a deletion in plakoglobin. *J Am Coll Cardiol* **38**, 1477-84.

Rebuzzini, P., Neri, T., Mazzini, G., Zuccotti, M., Redi, C. A. and Garagna, S. (2008). Karyotype analysis of the euploid cell population of a mouse embryonic stem

cell line revealed a high incidence of chromosome abnormalities that varied during culture. *Cytogenet Genome Res* **121**, 18-24.

Redkar, A., Montgomery, M. and Litvin, J. (2001). Fate map of early avian cardiac progenitor cells. *Development* **128**, 2269-79.

Riley, P. R., Gertsenstein, M., Dawson, K. and Cross, J. C. (2000). Early exclusion of hand1-deficient cells from distinct regions of the left ventricular myocardium in chimeric mouse embryos. *Dev Biol* **227**, 156-68.

Risau, W., Sariola, H., Zerwes, H. G., Sasse, J., Ekblom, P., Kemler, R. and Doetschman, T. (1988). Vasculogenesis and angiogenesis in embryonic-stem-cell-derived embryoid bodies. *Development* **102**, 471-8.

Robbins, J., Gulick, J., Sanchez, A., Howles, P. and Doetschman, T. (1990). Mouse embryonic stem cells express the cardiac myosin heavy chain genes during development in vitro. *J Biol Chem* **265**, 11905-9.

Roberts, T. M. and King, K. L. (1991). Centripetal flow and directed reassembly of the major sperm protein (MSP) cytoskeleton in the amoeboid sperm of the nematode, Ascaris suum. *Cell Motil Cytoskeleton* **20**, 228-41.

Roberts, T. M. and Stewart, M. (2000). Acting like actin. The dynamics of the nematode major sperm protein (msp) cytoskeleton indicate a push-pull mechanism for amoeboid cell motility. *J Cell Biol* **149**, 7-12.

Robertson, E. J. (1987). Teratocarcinomas and embryonic stem cells : a practical approach. Oxford: IRL.

Roitt, I. M. and Delves, P. J. (2001). Roitt's essential immunology. Oxford: Blackwell Science.

Ruiz, P., Brinkmann, V., Ledermann, B., Behrend, M., Grund, C., Thalhammer, C., Vogel, F., Birchmeier, C., Gunthert, U., Franke, W. W. et al. (1996). Targeted mutation of plakoglobin in mice reveals essential functions of desmosomes in the embryonic heart. *J Cell Biol* **135**, 215-25.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning : a laboratory manual. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory.

Sauer, B. and Henderson, N. (1989). Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res* **17**, 147-61.

Sauer, B. and McDermott, J. (2004). DNA recombination with a heterospecific Cre homolog identified from comparison of the pac-c1 regions of P1-related phages. *Nucleic Acids Res* **32**, 6086-95.

Schneider, A. G., Sultan, K. R. and Pette, D. (1999). Muscle LIM protein: expressed in slow muscle and induced in fast muscle by enhanced contractile activity. *Am J Physiol* **276**, C900-6.

Schneider, V. A. and Mercola, M. (2001). Wnt antagonism initiates cardiogenesis in Xenopus laevis. *Genes Dev* **15**, 304-15.

Schnutgen, F., De-Zolt, S., Van Sloun, P., Hollatz, M., Floss, T., Hansen, J., Altschmied, J., Seisenberger, C., Ghyselinck, N. B., Ruiz, P. et al. (2005). Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci U S A* **102**, 7221-6.

Schultheiss, T. M., Burch, J. B. and Lassar, A. B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes Dev* **11**, 451-62.

Scott, A. L., Dinman, J., Sussman, D. J. and Ward, S. (1989). Major sperm protein and actin genes in free-living and parasitic nematodes. *Parasitology* **98 Pt 3**, 471-8.

Sergueev, K., Yu, D., Austin, S. and Court, D. (2001). Cell toxicity caused by products of the p(L) operon of bacteriophage lambda. *Gene* **272**, 227-35.

Shepherd, P. and Dean, C. (2000). Monoclonal antibodies : a practical approach. Oxford: Oxford University Press.

Shirai, M., Miyashita, A., Ishii, N., Itoh, Y., Satokata, I., Watanabe, Y. G. and Kuwano, R. (1996). A gene trap strategy for identifying the gene expressed in the embryonic nervous system. *Zoolog Sci* **13**, 277-83.

Skarnes, W. C., Auerbach, B. A. and Joyner, A. L. (1992). A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev* **6**, 903-18.

Skarnes, W. C., von Melchner, H., Wurst, W., Hicks, G., Nord, A. S., Cox, T., Young, S. G., Ruiz, P., Soriano, P., Tessier-Lavigne, M. et al. (2004). A public gene trap resource for mouse functional genomics. *Nat Genet* **36**, 543-4.

Skehel, P. A., Armitage, B. A., Bartsch, D., Hu, Y., Kaang, B. K., Siegelbaum, S. A., Kandel, E. R. and Martin, K. C. (1995). Proteins functioning in synaptic transmission at the sensory to motor synapse of Aplysia. *Neuropharmacology* **34**, 1379-85.

Skehel, P. A., Fabian-Fine, R. and Kandel, E. R. (2000). Mouse VAP33 is associated with the endoplasmic reticulum and microtubules. *Proc Natl Acad Sci U S A* **97**, 1101-6.

Small, E. M. and Krieg, P. A. (2004). Molecular regulation of cardiac chamber-specific gene expression. *Trends Cardiovasc Med* **14**, 13-8.

Small, J. V., Furst, D. O. and Thornell, L. E. (1992). The cytoskeletal lattice of muscle cells. *Eur J Biochem* **208**, 559-72.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M. and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336**, 688-90.

Sohal, D. S., Nghiem, M., Crackower, M. A., Witt, S. A., Kimball, T. R., Tymitz, K. M., Penninger, J. M. and Molkentin, J. D. (2001). Temporally regulated and tissue-specific gene manipulations in the adult and embryonic heart using a tamoxifen-inducible Cre protein. *Circ Res* **89**, 20-5.

Srivastava, D. (2006). Genetic regulation of cardiogenesis and congenital heart disease. *Annu Rev Pathol* **1**, 199-213.

Srivastava, D., Cserjesi, P. and Olson, E. N. (1995). A subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270**, 1995-9.

Srivastava, D. and Olson, E. N. (2000). A genetic blueprint for cardiac development. *Nature* **407**, 221-6.

Srivastava, D., Thomas, T., Lin, Q., Kirby, M. L., Brown, D. and Olson, E. N. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nat Genet* **16**, 154-60.

Stahl, F. W. (1998). Recombination in phage lambda: one geneticist's historical perspective. *Gene* **223**, 95-102.

Stainier, D. Y. (2001). Zebrafish genetics and vertebrate heart formation. *Nat Rev Genet* **2**, 39-48.

Stanford, W. L., Caruana, G., Vallis, K. A., Inamdar, M., Hidaka, M., Bautch, V. L. and Bernstein, A. (1998). Expression trapping: identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. *Blood* **92**, 4622-31.

Stanford, W. L., Cohn, J. B. and Cordes, S. P. (2001). Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* **2**, 756-68.

Strubing, C., Ahnert-Hilger, G., Shan, J., Wiedenmann, B., Hescheler, J. and Wobus, A. M. (1995). Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature inhibitory and excitatory neurons. *Mech Dev* **53**, 275-87.

Sugawara, A., Goto, K., Sotomaru, Y., Sofuni, T. and Ito, T. (2006). Current status of chromosomal abnormalities in mouse embryonic stem cell lines used in Japan. *Comp Med* **56**, 31-4.

Sun, X., Meyers, E. N., Lewandoski, M. and Martin, G. R. (1999). Targeted disruption of Fgf8 causes failure of cell migration in the gastrulating mouse embryo. *Genes Dev* **13**, 1834-46.

Swaminathan, S., Ellis, H. M., Waters, L. S., Yu, D., Lee, E. C., Court, D. L. and Sharan, S. K. (2001). Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29**, 14-21.

Takahashi, N. and Kobayashi, I. (1990). Evidence for the double-strand break repair model of bacteriophage lambda recombination. *Proc Natl Acad Sci U S A* **87**, 2790-4.

Takeda, J., Keng, V. W. and Horie, K. (2007). Germline mutagenesis mediated by Sleeping Beauty transposon system in mice. *Genome Biol* **8 Suppl 1**, S14.

Tarr, D. E. and Scott, A. L. (2004). MSP domain proteins show enhanced expression in male germ line cells. *Mol Biochem Parasitol* **137**, 87-98.

Tarr, D. E. and Scott, A. L. (2005). MSP domain proteins. *Trends Parasitol* **21**, 224-31.

te Riele, H., Maandag, E. R. and Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells through homologous recombination with isogenic DNA constructs. *Proc Natl Acad Sci U S A* **89**, 5128-32.

Thomas, K. R. and Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-12.

Thomas, K. R., Folger, K. R. and Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44**, 419-28.

Thornell, L., Carlsson, L., Li, Z., Mericskay, M. and Paulin, D. (1997). Null mutation in the desmin gene gives rise to a cardiomyopathy. *J Mol Cell Cardiol* **29**, 2107-24.

Thyagarajan, B., Olivares, E. C., Hollis, R. P., Ginsburg, D. S. and Calos, M. P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* **21**, 3926-34.

Valancius, V. and Smithies, O. (1991). Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol Cell Biol* **11**, 1402-8.

Valenzuela, D. M., Murphy, A. J., Frendewey, D., Gale, N. W., Economides, A. N., Auerbach, W., Poueymirou, W. T., Adams, N. C., Rojas, J., Yasenchak, J. et al. (2003). High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21**, 652-9.

van Tintelen, J. P., Entius, M. M., Bhuiyan, Z. A., Jongbloed, R., Wiesfeld, A. C., Wilde, A. A., van der Smagt, J., Boven, L. G., Mannens, M. M., van Langen, I. M. et al. (2006). Plakophilin-2 mutations are the major determinant of familial arrhythmogenic right ventricular dysplasia/cardiomyopathy. *Circulation* **113**, 1650-8.

Ventura, A., Meissner, A., Dillon, C. P., McManus, M., Sharp, P. A., Van Parijs, L., Jaenisch, R. and Jacks, T. (2004). Cre-lox-regulated conditional RNA interference from transgenes. *Proc Natl Acad Sci U S A* **101**, 10380-5.

von Both, I., Silvestri, C., Erdemir, T., Lickert, H., Walls, J. R., Henkelman, R. M., Rossant, J., Harvey, R. P., Attisano, L. and Wrana, J. L. (2004). Foxh1 is essential for development of the anterior heart field. *Dev Cell* **7**, 331-45.

von Melchner, H., DeGregori, J. V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H. E. (1992). Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev* **6**, 919-27.

Voss, A. K., Thomas, T. and Gruss, P. (1998). Efficiency assessment of the gene trap approach. *Dev Dyn* **212**, 171-80.

Wackernagel, W. (1973). Genetic transformation in E. coli: the inhibitory role of the recBC DNase. *Biochem Biophys Res Commun* **51**, 306-11.

Waldo, K. L., Kumiski, D. H., Wallis, K. T., Stadt, H. A., Hutson, M. R., Platt, D. H. and Kirby, M. L. (2001). Conotruncal myocardium arises from a secondary heart field. *Development* **128**, 3179-88.

Wang, Z., Zhai, W., Richardson, J. A., Olson, E. N., Meneses, J. J., Firpo, M. T., Kang, C., Skarnes, W. C. and Tjian, R. (2004). Polybromo protein BAF180 functions in mammalian cardiac chamber maturation. *Genes Dev* **18**, 3106-16.

Ware, C. B., Siverts, L. A., Nelson, A. M., Morton, J. F. and Ladiges, W. C. (2003). Utility of a C57BL/6 ES line versus 129 ES lines for targeted mutations in mice. *Transgenic Res* **12**, 743-6.

Weitzer, G., Milner, D. J., Kim, J. U., Bradley, A. and Capetanaki, Y. (1995). Cytoskeletal control of myogenesis: a desmin null mutation blocks the myogenic pathway during embryonic stem cell differentiation. *Dev Biol* **172**, 422-39.

Westfall, M. V., Pasyk, K. A., Yule, D. I., Samuelson, L. C. and Metzger, J. M. (1997). Ultrastructure and cell-cell coupling of cardiac myocytes differentiating in embryonic stem cell cultures. *Cell Motil Cytoskeleton* **36**, 43-54.

Wobus, A. M., Kaomei, G., Shan, J., Wellner, M. C., Rohwedel, J., Ji, G., Fleischmann, B., Katus, H. A., Hescheler, J. and Franz, W. M. (1997). Retinoic acid accelerates embryonic stem cell-derived cardiac differentiation and enhances development of ventricular cardiomyocytes. *J Mol Cell Cardiol* **29**, 1525-39.

Xia, H., Winokur, S. T., Kuo, W. L., Altherr, M. R. and Bredt, D. S. (1997). Actinin-associated LIM protein: identification of a domain interaction between PDZ and spectrin-like repeat motifs. *J Cell Biol* **139**, 507-15.

Xu, X., Meiler, S. E., Zhong, T. P., Mohideen, M., Crossley, D. A., Burggren, W. W. and Fishman, M. C. (2002). Cardiomyopathy in zebrafish due to mutation in an alternatively spliced exon of titin. *Nat Genet* **30**, 205-9.

Yagi, T., Ikawa, Y., Yoshida, K., Shigetani, Y., Takeda, N., Mabuchi, I., Yamamoto, T. and Aizawa, S. (1990). Homologous recombination at c-fyn locus of mouse embryonic stem cells with use of diphtheria toxin A-fragment gene in negative selection. *Proc Natl Acad Sci U S A* **87**, 9918-22.

Yamagishi, H., Yamagishi, C., Nakagawa, O., Harvey, R. P., Olson, E. N. and Srivastava, D. (2001). The combinatorial activities of Nkx2.5 and dHAND are essential for cardiac ventricle formation. *Dev Biol* **239**, 190-203.

Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M. and Nakao, K. (2000). Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* **408**, 92-6.

Yanagawa, Y., Kobayashi, T., Ohnishi, M., Tamura, S., Tsuzuki, T., Sanbo, M., Yagi, T., Tashiro, F. and Miyazaki, J. (1999). Enrichment and efficient screening of ES cells containing a targeted mutation: the use of DT-A gene with the polyadenylation signal as a negative selection maker. *Transgenic Res* **8**, 215-21.

Yang, X. W., Model, P. and Heintz, N. (1997). Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol* **15**, 859-65.

Yang, Y. and Seed, B. (2003). Site-specific gene targeting in mouse embryonic stem cells with intact bacterial artificial chromosomes. *Nat Biotechnol* **21**, 447-51.

Yang, Z. and Vatta, M. (2007). Danon disease as a cause of autophagic vacuolar myopathy. *Congenit Heart Dis* **2**, 404-9.

Yelon, D., Ticho, B., Halpern, M. E., Ruvinsky, I., Ho, R. K., Silver, L. M. and Stainier, D. Y. (2000). The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. *Development* **127**, 2573-82.

Yu, D., Ellis, H. M., Lee, E. C., Jenkins, N. A., Copeland, N. G. and Court, D. L. (2000). An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**, 5978-83.

Zhang, Y., Buchholz, F., Muyrers, J. P. and Stewart, A. F. (1998). A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20**, 123-8.

Zhang, Y., Muyrers, J. P., Testa, G. and Stewart, A. F. (2000). DNA cloning by homologous recombination in Escherichia coli. *Nat Biotechnol* **18**, 1314-7.

Zhang, Y., Riesterer, C., Ayrall, A. M., Sablitzky, F., Littlewood, T. D. and Reth, M. (1996). Inducible site-directed recombination in mouse embryonic stem cells. *Nucleic Acids Res* **24**, 543-8.

Zheng, Q. and Zhao, Y. (2007). The diverse biofunctions of LIM domain proteins: determined by subcellular localization and protein-protein interaction. *Biol Cell* **99**, 489-502.

Zhou, J., Qu, J., Yi, X. P., Graber, K., Huber, L., Wang, X., Gerdes, A. M. and Li, F. (2007). Upregulation of gamma-catenin compensates for the loss of beta-catenin in adult cardiomyocytes. *Am J Physiol Heart Circ Physiol* **292**, H270-6.